Dimensionality Reduction with Automatic Dimension Assignment for Distributed Estimation

Jun Fang and Hongbin Li
Department of Electrical and Computer Engineering
Stevens Institute of Technology, Hoboken, NJ 07030
Emails: {Jun.Fang, Hongbin.Li@stevens.edu}

***Abstract*— We consider distributed estimation of a random vector parameter by a wireless sensor network (WSN). To meet stringent power and bandwidth budgets in WSN, local data compression is performed at each sensor to reduce the number of messages sent to a fusion center (FC). Under the constraint of a given *total* number of messages, our problem is to jointly determine the number of messages sent by each senor (a.k.a. *dimension assignment*) and design the corresponding compression matrix. The problem is formulated as a constrained optimization problem that minimizes the estimation mean-square error (MSE) at the FC. We analyze the problem using a subspace projection technique, which yields an efficient iterative solution. Numerical results are presented to illustrate the effectiveness of the proposed algorithm.**

***Index Terms*— Distributed estimation, joint dimension assignment and compression, wireless sensor network (WSN).**

I. Introduction

The problem of distributed estimation in wireless sensor networks (WSNs) has been of significant interest over the past few years. Due to limited power and communication bandwidth, some previous works (e.g., [1]–[4]) consider distributed estimation using aggressively quantized versions of the original observations. In this setup, quantization becomes an integral part of the estimation process and is critical to the estimation performance. Another category of methods (e.g., [5]–[9]), not relying on the above low-rate quantization strategy, follow an optimal decentralized compression-estimation approach to reduce the transmission requirement. In these methods, the data dimensionality is reduced before each sensor sends its data to a fusion center (FC). Upon receiving the compressed data, the FC combines them according to some fusion criterion to obtain a final estimate. The crux of these techniques is to design the compression matrix so as to minimize the estimation mean-square error (MSE), which has been extensively investigated by [5]–[9] under different fusion criterions and noise correlation scenarios. These techniques, however, require knowledge of the compression dimension associated with each sensor *a priori*. For the inhomogeneous environments, sensors at different locations may have dissimilar observation qualities, and it is necessary to turn off sensors with low-quality observations or use an aggressive compression dimension for these sensors. In this case, the above methods are limited since they require to set the compression dimension for each sensor, which is a tricky problem in practice. Hence, a data compression solution with automatic dimension allocation is desirable.

In this paper, we study joint compression dimension allocation and linear compression design under a bandwidth constraint. The bandwidth constraint is measured by the total number of real-valued messages (each message is a one-dimensional unquantized data sample) sent to the FC, or, equivalently, the sum of the total compression dimensions. We develop an efficient iterative algorithm that provides us an effective solution to the joint design problem. The proposed algorithm, unlike [5]–[9], can jointly determine the compression dimension and the corresponding compression matrix associated with each sensor. It, therefore, offers not only more flexibility but also a performance advantage over existing methods for distributed estimation in inhomogeneous environments.

We adopt the following notations throughout this paper. The notation $[\cdot]^T$ stands for matrix transpose. $E[\cdot]$ represents the mathematical expectation. Also, we use $\text{tr}(\mathbf{X})$ to represent the trace operation of matrix $\mathbf{X}$, $R(\mathbf{X})$ and $N(\mathbf{X})$ to indicate the range (column) space and null space of matrix $\mathbf{X}$, respectively. The symbol $\mathbf{I}_n$ represents the identity matrix of size $n \times n$. $\mathbb{R}^{n\times m}$ denotes the set of $n \times m$ matrices with real entries.

II. Problem Formulation

Consider a WSN consisting of N spatially distributed sensors, each sensor makes a noisy observation of the unknown vector parameter $\boldsymbol{\theta} \in \mathbb{R}^{p\times 1}$ (e.g., [7]):

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

where $\mathbf{H}_n \in \mathbb{R}^{q_n\times p}$ is the known observation matrix defining the input/output relation, $\mathbf{x}_n \in \mathbb{R}^{q_n\times 1}$ and $\mathbf{w}_n \in \mathbb{R}^{q_n\times 1}$ denote vector observation and noise, respectively. The unknown parameter $\boldsymbol{\theta}$ and the noise $\{\mathbf{w}_n\}$ are assumed statistically independent of each other, with zero-mean and covariance matrices $\mathbf{R}_\theta \triangleq E[\boldsymbol{\theta}\boldsymbol{\theta}^T]$ and $\mathbf{R}_w \triangleq E[\mathbf{w}\mathbf{w}^T]$, respectively, where $\mathbf{w} \triangleq [\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_N^T]^T$. We assume that there is no inter-sensor communication and the channel links between the sensors and the FC are ideal, i.e. noiseless. We also assume the knowledge of the covariance matrices $\mathbf{R}_\theta$ and $\mathbf{R}_w$ at the FC. In practice, they can be estimated from the sensors measurements in the absence/presence of signal (e.g., [5]). Note that in this paper, for simplicity, we only confine ourselves to the linear data model (1) and ideal channel scenarios. Dimensionality reduction under complicated scenarios like nonlinear data model and non-ideal channel links was studied in [9].

Let $\mathbf{x} \triangleq [\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_N^T]^T$, and $\mathbf{H} \triangleq [\mathbf{H}_1^T, \mathbf{H}_2^T, \ldots, \mathbf{H}_N^T]^T$. We can rewrite (1) in a more compact form as

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}, \tag{2}$$

where $\mathbf{H} \in \mathbb{R}^{q \times p}$, $\mathbf{x} \in \mathbb{R}^{q \times 1}$, $\mathbf{w} \in \mathbb{R}^{q \times 1}$, and $q \triangleq \sum_{n=1}^{N} q_n$. If the FC has access to all sensors data, the linear unbiased minimum variance (LUMV) estimate for $\boldsymbol{\theta}$ is known as (e.g., [5], [7])

$$\hat{\boldsymbol{\theta}} = \mathbf{R}_{\theta x}\mathbf{R}_x^{-1}\mathbf{x}, \tag{3}$$

with the estimation covariance matrix of $\hat{\boldsymbol{\theta}}$ given by

$$E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] = \mathbf{R}_\theta - \mathbf{R}_{\theta x}\mathbf{R}_x^{-1}\mathbf{R}_{\theta x}^T, \tag{4}$$

where $\mathbf{R}_{\theta x} \triangleq E[\boldsymbol{\theta}\mathbf{x}^T] = \mathbf{R}_\theta\mathbf{H}^T$ and $\mathbf{R}_x \triangleq E[\mathbf{x}\mathbf{x}^T] = \mathbf{H}\mathbf{R}_\theta\mathbf{H}^T + \mathbf{R}_w$. Also, $\mathbf{R}_x$ is assumed to be a positive-definite matrix with arbitrary spatial correlation (in some works, e.g. [7], $\mathbf{R}_x$ can be positive semi-definite). Although (3) gives the best estimate, this scheme requires sending all sensors data to the FC, which may be impractical for WSNs with stringent power and bandwidth budgets. A feasible solution is to reduce the transmission requirement through local data dimensionality reduction (also called "compression"). Due to the information redundancy arising from spatial correlation, this scheme is able to bring significant bandwidth savings while providing an acceptable estimation accuracy.

Since the compression dimension associated with each sensor is unknown and has to be determined, the previous works [5]–[9] are no longer applicable. To cope with such a situation, we introduce the following compression strategy to accommodate the unknown compression dimensions

$$\mathbf{z} = \mathbf{S}\mathbf{x} = \mathbf{S}\mathbf{H}\boldsymbol{\theta} + \mathbf{S}\mathbf{w}, \tag{5}$$

where

$$\mathbf{S} \triangleq \begin{bmatrix} \mathbf{c}_{1,k_1}\mathbf{E}_{k_1} \\ \mathbf{c}_{2,k_2}\mathbf{E}_{k_2} \\ \vdots \\ \mathbf{c}_{l,k_l}\mathbf{E}_{k_l} \end{bmatrix}, \tag{6}$$

is an $l \times q$ full row rank compression matrix, and $l \ll q$ is the pre-specified total number of messages to be sent to the FC. $\mathbf{E}_{k_i} \in \mathbb{R}^{q_{k_i} \times q}$, a selection matrix used to select sensor k_i^{th}'s data, is a sub-matrix consisting of rows $\bar{r}_1$ thru $\bar{r}_2$ of $\mathbf{I}_q$, where $\bar{r}_1 \triangleq \sum_{n=1}^{k_i-1} q_n + 1$, $\bar{r}_2 \triangleq \sum_{n=1}^{k_i} q_n$. That is, $\mathbf{E}_{k_i}$ is given by

$$\mathbf{E}_{k_i} \triangleq \begin{bmatrix} \mathbf{0}_{q_{k_i} \times q_1} & \ldots & \mathbf{I}_{q_{k_i}} & \ldots & \mathbf{0}_{q_{k_i} \times q_N} \end{bmatrix}. \tag{7}$$

$\mathbf{c}_{i,k_i} \in \mathbb{R}^{1 \times q_{k_i}}$ is a row vector which is used to linearly compress the selected sensor's data into a message.

We see that (6) provides a flexible framework to model the compression matrix with unknown compression dimensions because every row of $\mathbf{S}$ is free to choose any sensor. If multiple rows of $\mathbf{S}$, say t_n rows, select the same sensor n, it is equivalent to reducing sensor n's data dimensionality q_n to t_n; if no row corresponds to a certain sensor, this sensor is not selected.

Using the compressed data $\mathbf{z}$, the LUMV estimate of $\boldsymbol{\theta}$ and its estimation covariance matrix are given as follows, respectively

$$\hat{\boldsymbol{\theta}} = \mathbf{R}_{\theta x}\mathbf{S}^T(\mathbf{S}\mathbf{R}_x\mathbf{S}^T)^{-1}\mathbf{S}\mathbf{x}, \tag{8}$$

$$E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] = \mathbf{R}_\theta - \mathbf{R}_{\theta x}\mathbf{S}^T(\mathbf{S}\mathbf{R}_x\mathbf{S}^T)^{-1}\mathbf{S}\mathbf{R}_{\theta x}^T. \tag{9}$$

Naturally, we may wish to find an optimal compression matrix $\mathbf{S}$ to minimize the estimation mean-square error (MSE). That is,

$$\max_{\mathbf{S}} \quad \text{tr}\left(\mathbf{R}_{\theta x}\mathbf{S}^T(\mathbf{S}\mathbf{R}_x\mathbf{S}^T)^{-1}\mathbf{S}\mathbf{R}_{\theta x}^T\right). \tag{10}$$

We next study the optimization problem (10).

III. Proposed Approach

Because of the structure shown in (6), the optimization problem (10) is equivalent to determining the sensor indices $\{k_i\}_{i=1}^l$ and the corresponding compression vectors $\{\mathbf{c}_{i,k_i}\}_{i=1}^l$. Joint searching over the l sensor indices and the corresponding compression vectors, however, is practically infeasible since it involves a complexity that grows exponential with l. An alternative way, like in [9], is to simplify the problem by reducing the number of optimization variables. Specifically, we study how to determine the k^{th} row of $\mathbf{S}$ when the remaining $(l-1)$ rows are fixed, through which we can develop an efficient iterative algorithm to search for an effective, albeit suboptimal, solution.

Let $\mathbf{R}_x = \mathbf{Q}\mathbf{Q}^T$, where $\mathbf{Q} \in \mathbb{R}^{q \times q}$ can be obtained through eigenvalue decomposition (EVD) or Cholesky factorization. We can rewrite the cost function in (10) as

$$\begin{aligned} &\text{tr}\left(\mathbf{R}_{\theta x}\mathbf{S}^T(\mathbf{S}\mathbf{R}_x\mathbf{S}^T)^{-1}\mathbf{S}\mathbf{R}_{\theta x}^T\right) \\ \overset{(a)}{=}&\text{tr}\left(\mathbf{R}_{\theta x}\mathbf{Q}^{-T}\mathbf{Q}_s^T(\mathbf{Q}_s\mathbf{Q}_s^T)^{-1}\mathbf{Q}_s\mathbf{Q}^{-1}\mathbf{R}_{\theta x}^T\right) \\ \overset{(b)}{=}&\text{tr}\left(\mathbf{R}_{\theta x}\mathbf{Q}^{-T}\mathcal{T}[\mathbf{Q}_s]^T\mathcal{T}[\mathbf{Q}_s]\mathbf{Q}^{-1}\mathbf{R}_{\theta x}^T\right) \\ \overset{(c)}{=}&\text{tr}\left(\mathcal{T}[\mathbf{Q}_s]\mathbf{G}\mathcal{T}[\mathbf{Q}_s]^T\right), \end{aligned} \tag{11}$$

where $\mathbf{Q}_s \triangleq \mathbf{S}\mathbf{Q}$ in (a); in (b), $\mathcal{T}[\mathbf{X}]$ is used to represent a matrix transformation that transforms the rows of the full row rank matrix $\mathbf{X}$ into an orthonomal basis, i.e. we find a nonsingular matrix $\mathbf{A}$ such that $\mathcal{T}[\mathbf{X}] \triangleq \mathbf{A}\mathbf{X}$ admits: $(\mathcal{T}[\mathbf{X}])(\mathcal{T}[\mathbf{X}])^T = \mathbf{I}$; (c) follows from the definition $\mathbf{G} \triangleq \mathbf{Q}^{-1}\mathbf{R}_{\theta x}^T\mathbf{R}_{\theta x}\mathbf{Q}^{-T}$. Without loss of generality, we discuss the determination of the first row of $\mathbf{S}$, supposing that its last $(l-1)$ rows are given. We write

$$\mathbf{Q}_s = \mathbf{S}\mathbf{Q} \triangleq \begin{bmatrix} \mathbf{Q}_{s,1} \\ \mathbf{Q}_{s,2} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]} \\ \mathbf{Q}_{s,2} \end{bmatrix}, \tag{12}$$

where $\mathbf{Q}_{s,1} = \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}$ and $\mathbf{Q}_{s,2}$ denote the first row and the last $(l-1)$ rows of $\mathbf{Q}_s$, respectively, $\mathbf{Q}_{[r_1:r_2]}$ is a sub-matrix consisting of rows r_1 thru r_2 of $\mathbf{Q}$, in which $r_1 = \sum_{n=1}^{k_1-1} q_n + 1$ and $r_2 = \sum_{n=1}^{k_1} q_n$. We note that r_1 and r_2 are functions of the sensor index k_1. Combining (10–12), hence the optimization becomes

$$\max_{k_1, \mathbf{c}_{1,k_1}} \text{tr}\left(\mathcal{T}\left[\begin{bmatrix} \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]} \\ \mathbf{Q}_{s,2} \end{bmatrix}\right]\mathbf{G}\mathcal{T}\left[\begin{bmatrix} \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]} \\ \mathbf{Q}_{s,2} \end{bmatrix}\right]^T\right). \tag{13}$$

Note that the matrix transformation $\mathcal{T}[\cdot]$ in (13) is not unique since if $\mathcal{T}[\mathbf{X}] = \mathbf{AX}$ satisfies $(\mathcal{T}[\mathbf{X}])(\mathcal{T}[\mathbf{X}])^T = \mathbf{I}$, then $\mathcal{T}[\mathbf{X}] = \mathbf{UAX}$ also works for any orthogonal matrix $\mathbf{U}$, and it makes no difference to the cost function. Our goal is to seek one transformation that decouples the optimization variable from the invariables. This can be accomplished by using a subspace projection technique as described next.

We construct the orthogonal projection onto $R(\mathbf{Q}_{s,2}^T)$ and $N(\mathbf{Q}_{s,2})$, respectively, as

$$\mathbf{P} = \mathbf{Q}_{s,2}^T(\mathbf{Q}_{s,2}\mathbf{Q}_{s,2}^T)^{-1}\mathbf{Q}_{s,2}, \tag{14}$$

$$\mathbf{P}^\perp = \mathbf{I}_q - \mathbf{Q}_{s,2}^T(\mathbf{Q}_{s,2}\mathbf{Q}_{s,2}^T)^{-1}\mathbf{Q}_{s,2}. \tag{15}$$

Then, we have the following result.

Proposition 1: For any full row rank matrix $\mathbf{Q}_s$ given in (12), its matrix transformation $\mathcal{T}[\mathbf{Q}_s]$ can be written as

$$\mathcal{T}\left[\begin{bmatrix} \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]} \\ \mathbf{Q}_{s,2} \end{bmatrix}\right] = \begin{bmatrix} \gamma_{k_1}(\mathbf{P}^\perp\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T)^T \\ \mathcal{T}[\mathbf{Q}_{s,2}] \end{bmatrix}, \tag{16}$$

where

$$\gamma_{k_1} \triangleq \frac{1}{\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}\mathbf{P}^\perp\mathbf{P}^\perp\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T}, \tag{17}$$

is a scalar normalizing the vector $(\mathbf{P}^\perp\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T)$, and $\mathcal{T}[\mathbf{Q}_{s,2}]$ represents any matrix transformation that transforms the rows of $\mathbf{Q}_{s,2}$ into an orthonormal basis.

Proof: See Appendix A. ■

Observe that (16) has successfully separated the optimization variable from the invariable as we desired. By utilizing Proposition 1, the optimization (13) can therefore be re-expressed as

$$\max_{k_1,\mathbf{c}_{1,k_1}} \quad \left(\gamma_{k_1}^2\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}\mathbf{P}^\perp\mathbf{G}\mathbf{P}^\perp\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T\right) + \text{tr}\left(\mathcal{T}[\mathbf{Q}_{s,2}]\mathbf{G}\mathcal{T}[\mathbf{Q}_{s,2}]^T\right), \tag{18}$$

where the second term is independent of k_1 and $\mathbf{c}_{1,k_1}$ and thus can be ignored. The above optimization can be further reduced to a one-dimensional search by replacing $\mathbf{c}_{1,k_1}$ with its optimum $\mathbf{c}_{1,k_1}^*$ for every possible k_1. Given a specified k_1, the optimum $\mathbf{c}_{1,k_1}^*$ is determined by

$$\max_{\mathbf{c}_{1,k_1}} \quad \left(\gamma_{k_1}^2\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}\mathbf{P}^\perp\mathbf{G}\mathbf{P}^\perp\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T\right). \tag{19}$$

We now discuss how to solve (19). Notice that for a specified k_1, $\mathbf{Q}_{[r_1:r_2]}$ is fixed, and $\gamma_{k_1}\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}\mathbf{P}^\perp$ is a unit-norm vector which is a linear combination of the rows of $\mathbf{\Lambda} \triangleq \mathbf{Q}_{[r_1:r_2]}\mathbf{P}^\perp$. Let $\mathbf{\Lambda} \triangleq \mathbf{UDV}^T$ denote the reduced singular value decomposition, where $\mathbf{U} \in \mathbb{R}^{q_{k_1}\times r}$, $\mathbf{D} \in \mathbb{R}^{r\times r}$, and $\mathbf{V} \in \mathbb{R}^{q\times r}$, $r \le q_{k_1}$ is the rank of $\mathbf{\Lambda}$. Thus we can write

$$\gamma_{k_1}\mathbf{c}_{1,k_1}\mathbf{\Lambda} = \gamma_{k_1}\mathbf{c}_{1,k_1}\mathbf{UDV}^T \triangleq \mathbf{cV}^T, \tag{20}$$

where $\mathbf{c} \triangleq \gamma_{k_1}\mathbf{c}_{1,k_1}\mathbf{UD}$ is an r-dimensional row vector of unit norm, i.e. $\|\mathbf{c}\| = 1$. Therefore (19) is equivalent to

$$\max_{\|\mathbf{c}\|=1} \quad \left(\mathbf{cV}^T\mathbf{G}\mathbf{V}\mathbf{c}^T\right), \tag{21}$$

where the row vector $\mathbf{c}$ can be obtained as the eigenvector associated with the largest eigenvalue. Let $\gamma_{k_1} = 1$, $\mathbf{c}_{1,k_1}^*$ can be easily solved from

$$\mathbf{c}_{1,k_1}^*\mathbf{UD} = \mathbf{c}. \tag{22}$$

(22) admits an unique exact solution when $r = q_{k_1}$ and numerous exact solutions when $r < q_{k_1}$. For the latter case, we can pick any one of the solutions. After obtaining $\mathbf{c}_{1,k_1}^*$ for each $k_1 \in \{1,\ldots,N\}$, k_1 is finally determined as

$$\max_{k_1} \quad \left(\mathbf{c}_{1,k_1}^*\mathbf{Q}_{[r_1:r_2]}\mathbf{P}^\perp\mathbf{G}\mathbf{P}^\perp\mathbf{Q}_{[r_1:r_2]}^T(\mathbf{c}_{1,k_1}^*)^T\right). \tag{23}$$

From the above discussion, we see that through the decoupling transform of Proposition 1, the optimization (13) can be solved via (21)–(23). This effectively establishes an iterative algorithm by successively optimizing and replacing each row of $\mathbf{S}$. The algorithm is summarized as follows

1) Randomly generate a selection matrix $\mathbf{S}^{(0)}$ as an initialization.
2) At iteration $i+1$ ($i = 0, 1, \ldots$): via (21)–(23), determine $\mathbf{S}_{[1]}^{(i+1)}$ given: $\{\mathbf{S}_{[2]}^{(i)},\ldots,\mathbf{S}_{[l]}^{(i)}\}$; determine $\mathbf{S}_{[k]}^{(i+1)}$ given: $\{\mathbf{S}_{[1]}^{(i+1)},\ldots,\mathbf{S}_{[k-1]}^{(i+1)},\mathbf{S}_{[k+1]}^{(i)},\ldots,\mathbf{S}_{[l]}^{(i)}\}$ for $k = 2,\ldots,l$. Here we use $\mathbf{S}_{[k]}$ to denote the k^{th} row of $\mathbf{S}$.
3) Go to Step 2 if $|f(\mathbf{S}^{(i+1)}) - f(\mathbf{S}^{(i)})| > \epsilon$, where $f(\cdot)$ denotes the cost function defined in (10), ϵ is a prescribed tolerance value; otherwise stop.

Clearly, in this algorithm, every iteration results in a non-decreasing cost function value. Although not guaranteed to converge to the global maximum, this algorithm converges to a stationary point and provides us a practical compression matrix design.

IV. Numerical Results

We present numerical results to illustrate the estimation performance of the proposed algorithm. In our simulations, we set $N = 5$, $p = 5$, and $q_n = 7$ for any $n \in \{1,\ldots,N\}$. The observation matrices $\{\mathbf{H}_n\}$ are randomly generated with its elements independently chosen as Gaussian random variables with zero mean and variance σ_s^2. To simulate an inhomogeneous environment with varying signal-to-noise ratio (SNR), we let $\sigma_s^2 = 0.1$ for three sensors and $\sigma_s^2 = 1$ for the rest of two sensors. Also, the signal and noise covariance matrices are chosen to be $\mathbf{R}_\theta = \mathbf{I}_p$ and $\mathbf{R}_w = 0.1\mathbf{I}_q$, respectively. Note that the sensors' observations are still spatially correlated with covariance matrix $\mathbf{R}_x = \mathbf{H}\mathbf{R}_\theta\mathbf{H}^T + \mathbf{R}_w$.

We compare our proposed method with the non-compression scheme and the method [8] which requires the compression dimensions to be set *a priori*. For the setup considered herein, the methods of [5], [9] yield similar performance as that of [8] and are thus omitted. The non-compression scheme uses all sensors data with its estimation covariance matrix given by (4), which provides a benchmark (lower bound) on the achievable performance of all rate-constrained methods. For [8], we examine the case where the compression dimensions assigned to all sensors are identical. Fig. 1 shows the mean-square error (MSE) of the three schemes as a function of the number of messages sent to the

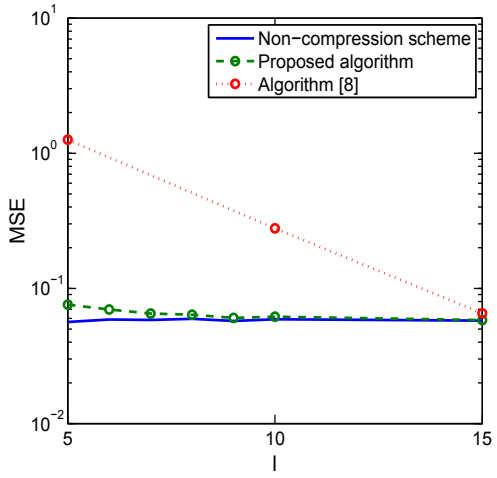

Fig. 1. MSEs versus the total number of messages l sent by all sensors.

FC, l. The results are averaged over 200 Monte Carlo runs, and the observation matrices $\{\mathbf{H}_n\}$ are independently generated for each run. In Fig. 1, the three points on the curve of the method [8] correspond to $(t_1,t_2,t_3,t_4,t_5) = (1,1,1,1,1)$, $(t_1,t_2,t_3,t_4,t_5) = (2,2,2,2,2)$, and $(t_1,t_2,t_3,t_4,t_5) = (3,3,3,3,3)$, respectively, where t_n denotes the compression dimension associated with sensor n. From Fig. 1, we see that our proposed method has a performance advantage over the method [8] with identical compression dimension assignment. The performance gain is primarily due to the fact that our scheme is able to make more efficient use of the total compression dimensions l by taking into account sensor disparity. In particular, it automatically assigns more dimensions to sensors with high-quality observations than to noisier sensors. For example, the dimension assignments obtained from one realization by our proposed method are as follows: $l = 5$: $(t_1,t_2,t_3,t_4,t_5) = (0,0,0,3,2)$, $l = 10$: $(t_1,t_2,t_3,t_4,t_5) = (0,0,0,5,5)$, $l = 15$: $(t_1,t_2,t_3,t_4,t_5) = (1,1,1,6,6)$. We also observe that, for a moderate l, our proposed algorithm attains an estimation accuracy comparable to that of the non-compression scheme. Notice that for the non-compression scheme, a total number of $q = 35$ messages are sent to the FC. Hence a considerable bandwidth savings is achieved by our proposed method.

V. Conclusion

The problem of distributed parameter estimation is studied in this paper. In order to meet the bandwidth constraint in wireless sensor networks, each sensor compresses its data before transmitting it to the fusion center. We developed an efficient iterative algorithm that jointly determine the compression dimension and the corresponding compression matrix associated with each sensor. Simulation results show that our proposed algorithm can effectively capture the observation quality difference across the sensors and provide efficient dimension assignment. Also, it can achieve a considerable bandwidth savings at a small performance degradation as compared with the non-compression scheme.

Appendix A
Proof of Proposition 1

To prove (16), we need to show that the rows of the matrix on the right hand side of (16) is an orthonormal basis for the row space of $\mathbf{Q}_s$. For notational convenience, let

$$\mathbf{T} \triangleq \begin{bmatrix} \gamma_{k_1}(\mathbf{P}^{\perp}\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T)^T \\ \mathcal{T}[\mathbf{Q}_{s,2}] \end{bmatrix}.$$

We, firstly, verify that the rows of $\mathbf{T}$ are normalized orthogonal vectors. Since $\mathbf{Q}_{s,2}\mathbf{P}^{\perp}\mathbf{b} = \mathbf{0}$ for any $\mathbf{b} \in \mathbb{R}^{q\times 1}$, we have $\mathbf{Q}_{s,2}\mathbf{P}^{\perp}\mathbf{Q}_{[r_1:r_2]}^T = \mathbf{0}$ and consequently $\mathcal{T}[\mathbf{Q}_{s,2}]\mathbf{P}^{\perp}\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T = \mathbf{0}$. Therefore we have

$$\mathbf{T}\mathbf{T}^T = \mathbf{I}_l. \tag{24}$$

We now show that $\mathbf{Q}_s$ and $\mathbf{T}$ have the same row space. To this end, we demonstrate that every row of $\mathbf{Q}_s$ can be represented by a linear combination of the rows of $\mathbf{T}$, and vice versa. Since we can write $\mathcal{T}[\mathbf{Q}_{s,2}] = \mathbf{A}\mathbf{Q}_{s,2}$, where $\mathbf{A} \in \mathbb{R}^{(l-1)\times(l-1)}$ is an invertible matrix, we only need to examine the first row of the respective matrices. We have

$$(\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]})^T = \mathbf{P}^{\perp}(\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]})^T + \mathbf{P}(\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]})^T, \tag{25}$$

and further we write

$$\begin{aligned}\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]} =& (\mathbf{P}^{\perp}\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T)^T + \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}\mathbf{P}^T \\ =& (\mathbf{P}^{\perp}\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T)^T \\ &+ \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}\mathbf{Q}_{s,2}^T(\mathbf{Q}_{s,2}\mathbf{Q}_{s,2}^T)^{-1}\mathbf{Q}_{s,2} \\ \overset{(a)}{=}& (\mathbf{P}^{\perp}\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T)^T + \mathbf{a}^T\mathcal{T}[\mathbf{Q}_{s,2}],\end{aligned} \tag{26}$$

where $\mathbf{a}^T \triangleq \mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}\mathbf{Q}_{s,2}^T(\mathbf{Q}_{s,2}\mathbf{Q}_{s,2}^T)^{-1}\mathbf{A}^{-1}$ in (a). From (26) we observe that $\mathbf{c}_{1,k_1}\mathbf{Q}_{[r_1:r_2]}$ and $(\mathbf{P}^{\perp}\mathbf{Q}_{[r_1:r_2]}^T\mathbf{c}_{1,k_1}^T)^T$ are linear combinations of the rows of the matrices $\mathbf{T}$ and $\mathbf{Q}_s$, respectively. The proof is completed here.

References

[1] 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.

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

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

[4] 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.

[5] K. Zhang, X. R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion – Part VI: sensor data compression," in *Proc. of the Intl. Conf. on Info. Fusion*, Queensland, Australia, 2003.

[6] Z.-Q. Luo, G. B. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," in *Proc. of the Intl. Symp. on Info. Theory*, Adelaide, Australia, Sept. 2005.

[7] 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.

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

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