Energy-Efficient Estimation of Correlated Data in Wireless Sensor Networks

Israfil Bahceci and Amir K. Khandani
Department of Elect. & Comp. Eng.
University of Waterloo
Waterloo, ON N2L3G1, CA
Email: {bahceci,khandani}@shannon2.uwaterloo.ca

***Abstract*—In this paper, we study the energy-efficient distributed estimation problem for a wireless sensor network where a physical phenomena that produces correlated data is sensed by a set of spatially distributed sensor nodes and the resulting noisy observations are transmitted to a fusion center via noise-corrupted channels. We assume a Gaussian network model where (i) the data being sensed at different sensors are correlated and the correlation structure (in the form of a correlation matrix) is known at the fusion center, (ii) the links between the local sensors and the fusion center are subject to multipath fading plus AWGN, and the fading gains are available to the receiver node, and (iii) the central node uses the squared error distortion metric. We first determine the optimum power-distortion regions assuming (i) a multiple-letter, and (ii) a single-letter square distortion characterization. Next, for the two distortion characterization, we investigate the performance of an uncoded transmission approach where the noisy observations are only amplified-and-forwarded to the fusion center. At the fusion center, two different estimators are considered: (i) minimum mean-square error estimator (MMSE) that exploits the correlation, and (ii) best linear unbiased estimator (BLUE) that does not require or exploit the knowledge of the correlation matrix. For both estimators, we solve for the optimal power allocation that results in a minimum total transmission power while satisfying some distortion level for the estimate (for both multiple-letter and single-letter distortion metrics). The numerical comparisons between the two schemes indicate that the MMSE requires less power to attain the same distortion provided by the BLUE. Furthermore, comparisons between power-distortion region achieved by the theoretically optimum system and the uncoded system indicates that the performance gap between the two system becomes small for low level of correlation between the sensor observations. If observations at all sensor nodes are uncorrelated, the uncoded system attains the theoretically optimum system performance.**

I. Introduction

Wireless sensor networks (WSNs) enhance the system capabilities in many application areas including environment monitoring, health, surveillance, etc. [1]. Networks of sensor systems allow for many distributed processing and cooperative communication techniques. In this paper, we focus on an estimation problem where each sensor sends its observation to a fusion center where a global estimation is made. Because of the hard energy limitations, a significant research problem is to develop schemes that minimize the transmission energy while satisfying a certain distortion level.

Various approaches can be followed to solve this problem. On the one hand, one can digitize all observed data at the local sensors using distributed compression/coding algorithms and transmit the digitized data to the fusion center. This is mainly suggested by the aforementioned source-channel separation theorem of Shannon [2]. With this view, this problem falls into the multiterminal source coding problem where the main issue is to characterize the rate-distortion region, i.e., to determine all the rate vectors $(R_1, \ldots, R_K)$ at which the source samples at all K sensors can be encoded separately and then decoded jointly attaining a prespecified distortion level $(D_1, \ldots, D_K)$. Coding for multiterminal source-channel communications and the associated rate-distortion region is extensively studied in the literature and several partial solutions are reported to date [3-7]. In [4], Slepian and Wolf consider lossless coding of discrete memoryless correlated sources and show that it is possible to attain the rate-distortion region of joint encoding/decoding by the scheme where the sources are encoded separately while decoded jointly. In [5], Wyner and Ziv initiate the research on lossy source coding where they determine the rate-distortion region for source coding with side information. While the optimal solution for a general setting is not known yet, for some of the special cases, e.g., [7], it is shown that separate encoding/joint decoding might incur some rate loss. An important class of the multi-terminal source/channel coding problems is the Chief Executive Officer (CEO) problem where a set of separate agents make noisy observations of a common source and transmit a summary (encoded version) to the CEO where the final decision (e.g., estimation) is made [8–12].

The separate source-channel coding generally incurs large delays and computation complexity. However, in some cases, such as for a point-to-point communcation over an additive white Gaussian noise channel, it is possible use a simple amplify-and-forward approach [13-15] to achieve optimality. For multiterminal source-channel communications, however, it is not clear weather this approach performs better [12], [16]. Nevertheless, since the information in sensor networks is in general delay sensitive, and because of the bandwidth and hard energy constraints in sensor networks, one requires simple coding/processing methods, rather than relying on the source-channel separation theorem. In [17], Cui et al. study a simple uncoded analog transmission method for data gathering in sensor network where it is assumed that the noise corrupted version of the same signal is observed each local node and a single-side band (SSB) analog transmission is employed to transmit the real-valued observations to the fusion center. This analog approach is very simple since it solely relies on an amplify-and-forward technique. A similar approach is also studied in [16] by Gastpar and Vetterli and it is shown that the uncoded scheme results in a larger decay of the distortion in the estimation, and as the number of sensor nodes goes to infinity, this system attains the best estimation performance.

In sensor networks, the sensor observations are very likely

1-4244-0350-2/06/$20.00 ©2006 IEEE

to vary from one sensor to the other, but generally the observations might be strongly correlated, especially for a dense sensor network [18]. Therefore, in this paper, we consider a network where the sensor nodes observe spatially correlated data and transmit their summary to a fusion center for the purpose of distributed estimation. This problem is similar to the Quadratic Gaussian CEO problem e.g., [9], with the exception that instead of the assumption that the same signal is observed by each agent, we assume that the agents (sensor nodes) are exposed to a correlated data field. For this problem, we first derive the rate-distorion and power-disrotion region that can be attained by the theoretically optimum system. Next, we study the performance of the simple uncoded transmission based on amplify-and-forward transmission. In fact, such a transmission approach can be imagined as a *rate-1 joint source-channel coding with separate encoders* at different sensor nodes and the estimation can be viewed as a *joint decoding* at the fusion center. Here, we study the performance of two different estimators: (i) minimum mean-square estimator (MMSE), requiring and exploiting the knowledge of the correlation in the observed field and (ii) minimum (best) linear unbiased estimator (BLUE), free from any statistical information; and in both cases, we consider two distortion measures: (i) the mean-squared error distortion averaged across the sensor nodes as a single-letter characterization of the distortion, and (ii) the individual mean-squared error distortion for the estimation quality of the signal from each sensor node as a multiple-letter characterization of the distortion. The BLUE considered in [17] is different from the one being considered here since here we study the estimation of a spatially correlated field. As we shall see in Section III, the solutions to these two problems differ significantly from eachother. The numerical optimization results indicate that the MMSE outperforms the BLUE in terms of energy efficiency. Furthermore, performance comparions between the theoretically optimum system and the uncoded system with MMSE at the fusion center indicate that for the case of low level correlation among sensor observations, the uncoded transmission performance gets close to the optimum theoretical system performance.

The organization of the paper is as follows: In the next section, we describe the sensor network model and specify the parameters for analog transmission. In Section III, we briefly summarize the estimator performance and derive the rate- and power-disrotion regions for the theoretically optimum system performance using the Shannon bounds. In this section, we investigate the energy efficient distributed estimation problem for the uncoded schemes employing the minimum mean-square error estimator and the best linear unbiased estimator, formulate the power optimization problems and solve them. Section IV presents numerical results for the power allocation problem, and finally Section V summarizes the results and future directions.

II. System Model

Assume that there are K sensors and the observation at the k^{th} sensor at time t, $x_k(t)$, is a random signal given by

$$x_k(t) = \theta_k(t) + n_k(t),\ t = 1, 2, \ldots \tag{1}$$

where $\theta_k(t)$ is the value of the observed field and $n_k(t)$ is the additive white Gaussian (AWGN) noise with variance σ_k^2. Let $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_k(t)]$ and $\mathbf{n} = [n_1(t), \ldots, n_K(t)]$. We assume that $\boldsymbol{\theta}(t)$ is an independently and identically distributed Gaussian vector whose autocorrelation is given by

$$\mathbf{R}_{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\boldsymbol{\theta}^H\}.$$

The transmitted signal from sensor node k is given by

$$y_k(t) = \sqrt{\alpha_k}\theta_k(t)$$

where α_k is power scaling paramater. Thus, the average transmission power is given by $\alpha_k P_{x_k}$ where $P_{x_k} = [\mathbf{R}_{\boldsymbol{\theta}}]_{k,k} + \sigma_k^2$. Assuming a flat fading channel with a gain factor of g_k for the k^{th} node, we can express the received signal as

$$r_k(t) = \sqrt{\alpha_k g_k}\theta_k(t) + \sqrt{\alpha_k g_k} n_k(t) + n_{c_k}(t)$$

with $n_{c_k}(t)$ denoting the channel noise of variance ξ_k^2. In vector form, we have the input-output relation (for brevity, we drop the time parameter)

$$\mathbf{r} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v} \tag{2}$$

where

$$\begin{aligned} \mathbf{r} &= [r_1, \ldots, r_K]^T \text{ and } \boldsymbol{\theta} = [\theta_1, \ldots, \theta_K], & (3)\\ \mathbf{H} &= \text{diag}(\sqrt{\alpha_1 g_1}, \ldots, \sqrt{\alpha_K g_K}), & (4)\\ \mathbf{v} &= \text{diag}(\sqrt{\alpha_1 g_1} n_1 + n_{c1}, \ldots, \sqrt{\alpha_K g_K} + n_{cK}). & (5) \end{aligned}$$

Assuming independently and identically distributed additive observation noise and the channel noise, we have the noise covariance matrix

$$\mathbf{R}_{\mathbf{v}} = \text{diag}(\alpha_1 g_1 \sigma_1^2 + \xi_1^2, \ldots, \alpha_K g_K \sigma_K^2 + \xi_K^2) \tag{6}$$

where diag$(\cdot)$ denotes a diagonal matrix formed from its vector argument.

Let $\hat{\boldsymbol{\theta}} = f(\boldsymbol{\theta})$ denote any estimate of $\boldsymbol{\theta}$. The error covariance matrix is defined by $\mathbf{R}_{\boldsymbol{\epsilon}} = E\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\}$. The Cramer-Rao bound on $\mathbf{R}_{\boldsymbol{\epsilon}}$ for the signal model in (2) is given by [19]

$$\mathbf{R}_{\boldsymbol{\epsilon}} \geq (\mathbf{H}^T \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{H} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1}, \tag{7}$$

which can be attained by the linear minimum mean square error estimator (MMSE) for a Gaussian signal model. Note that the k^{th} diagonal entry of the error covariance matrix, $[\mathbf{R}_{\boldsymbol{\epsilon}}]_{kk}$, is the squared error distortion at node k. In this paper, we consider two distortion characterization: (i) multiple-letter distortion metric

$$\mathbf{D} = [D_1, \ldots, D_K] \tag{8}$$

where $D_k = E\{|\theta_k - \hat{\theta_k}|^2\} = [\mathbf{R}_{\boldsymbol{\epsilon}}]_{kk}$, and (ii) a single-letter distortion metric

$$\begin{aligned} D_0 &= \frac{1}{K}\text{tr}(E\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\}) & (9)\\ &= \frac{1}{K}\text{tr}(\mathbf{R}_{\boldsymbol{\epsilon}}). & (10) \end{aligned}$$

This latter distortion is a measure of average mean squared-error in the estimation across all sensor nodes.

III. Energy-Efficient Estimation

In this section, we first derive the power-distortion region for the system described above. Then, we study the performance of uncoded transmission and solve the optimum power allocation problem with MMSE and BLUE.

A. Shannon Bounds and Power-Distortion Region

The Cramer-Rao bound in (7) specify the best error-covariance matrix attainable by any estimator for the prescribed signal model in (2). The fundamental bounds on the estimation quality can be derived by using the Shannon rate-distortion and capacity formulas, which are originally derived to assess the reconstruction quality for the source coding with a fidelity criterion, and the ultimate transmission rate over a given channel [14]. The rate-distortion region for the problem at hand can be characterized by using a Gaussian test channel [12], [10], [20]. First, we derive the $\mathcal{P}-\mathcal{D}$-region for multiple-letter distortion. The region to the single-letter distortion will follow similarly. Let $\mathbf{Z}_k = \mathbf{X}_k + \mathbf{n}_{c,k}$ where $\mathbf{n}_{c,k}$ is a zero mean Gaussian noise vector with covariance matrix $\mathbf{R}_\nu = \mathrm{diag}(\eta_1^2, \ldots, \eta_K^2)$ and $\mathbf{X}_k = \boldsymbol{\theta}_k + \mathbf{n}_k$ [12]. The set of variances η_k, $k = 1, \ldots, K$ for which the mean-squared error distortion at node k, $k = 1, \ldots, K$, is less than D_k is given by

$$\Psi_M(\mathbf{D}) \triangleq \left\{(\eta_1^2, \ldots, \eta_K^2) : D_k \geq \left[\left((\mathbf{R}_\nu + \mathbf{R_n})^{-1} + \mathbf{R}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\right]_{kk}\right\}. \tag{11}$$

Then, the rate-distortion region follows as

$$\mathcal{R}(\mathbf{D}) = \bigcup_{(\eta_1^2,\ldots,\eta_K^2)\in\Psi_M(\mathbf{D})} \mathcal{R}(\mathbf{D}; \eta_1^2, \ldots, \eta_K^2) \tag{12}$$

where

$$\mathcal{R}(\mathbf{D}; \eta_1^2, \ldots, \eta_K^2) = \left\{(R_1, \ldots, R_K) : \sum_{i\in\mathcal{A}} R_i \geq I(X_{\mathcal{A}}; Z_{\mathcal{A}} | Z_{\mathcal{A}^c})\right\}$$

for $\forall \mathcal{A} \subseteq \mathcal{I}_K$, $\mathcal{I}_K = \{1, \ldots, K\}$, and where $X_{\mathcal{A}}$ and $I(\cdot)$ denote the set $\{X_k : k \in \mathcal{A}\}$ and the mutual information, respectively. Combining this region with the Shannon channel capacity for each sensor node (assuming a zero-mean unit-variance additive white Gaussian channel noise), $R_k \leq \frac{1}{2}\log(1+P_k g_k)$, we finally arrive at the power-distortion region

$$\mathcal{P}(\mathbf{D}) = \bigcup_{(\eta_1^2,\ldots,\eta_K^2)\in\Psi_M(\mathbf{D})} \mathcal{P}(\mathbf{D}; \eta_1^2, \ldots, \eta_K^2) \tag{13}$$

where

$$\mathcal{P}(\mathbf{D}; \eta_1^2, \ldots, \eta_K^2) = \left\{(P_1, \ldots, P_K) : \prod_{i\in\mathcal{A}} (1 + P_k g_k) \geq \left(\frac{|\mathbf{R}_{X_{\mathcal{A}} Z_{\mathcal{A}^c}}||\mathbf{R}_Z|}{|\mathbf{R}_{X_{\mathcal{A}} Z}||\mathbf{R}_{Z_{\mathcal{A}^c}}|}\right)^\kappa\right\} \tag{14}$$

for $\forall \mathcal{A} \subseteq \mathcal{I}_K$ where κ is the source/channel code rate (e.g., for uncoded transmission, $\kappa = 1$), $X_{\mathcal{A}_1} Z_{\mathcal{A}_2}$ denotes the vector formed by stacking the elements of $X_{\mathcal{A}_1}$ and $Z_{\mathcal{A}_2}$, and $\mathbf{R}_{X_{\mathcal{A}_1} Z_{\mathcal{A}_2}}$ denotes the corresponding covariance matrix. We note that the determinants in (14) can be evaluated using the determinant formula for the block matrices after reordering the terms in the matrices with suitable permutation matrices corresponding to the subsets $\mathcal{A}$ and $\mathcal{A}^c$.

Because the region $\mathcal{P}(\mathbf{D})$ is convex, the power allocation problem to achieve minimum transmission power can be cast as

$$\begin{array}{ll} \min & \sum_{k=1}^K P_k \\ \text{s.t.} & (P_1, \ldots, P_K) \in \mathcal{P}(\mathbf{D}). \end{array} \tag{15}$$

Note that the solution to this problem provides the power vector that can attain the theoretical minimum transmission power (within all possible coding/processing schemes) while satisfying a prescribed distortion level $\mathbf{D}$.

The analysis for the single-letter distortion metric is very similar to the multiple-letter case above. We only need to replace the region $\Psi_M(\mathbf{D})$ in (11) by

$$\Psi_S(D_0) \triangleq \left\{(\eta_1^2, \ldots, \eta_K^2) : D_0 \geq \frac{1}{K}\mathrm{tr}\left((\mathbf{R}_\nu + \mathbf{R_n})^{-1} + \mathbf{R}_{\boldsymbol{\theta}}^{-1}\right)^{-1}\right\}.$$

and $\mathbf{D}$ by D_0 in all subsequent equations (12)-(15)

B. Minimum Mean-Square Estimation and Optimal Power Allocation

1) Individual Distortion Constraint: First, we assume that the estimation error for the sample observed at node k is constrained to be no more than D_k. With the knowledge of the correlation matrix $R_{\boldsymbol{\theta}}$, the MMSE for $\boldsymbol{\theta}$ in (1) is given by

$$\hat{\boldsymbol{\theta}} = \mathbf{R}_{\boldsymbol{\theta}}\mathbf{H}^T(\mathbf{H}\mathbf{R}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{R_v})^{-1}\mathbf{y} \tag{16}$$

and the minimum mean-squared error covariance matrix is given by

$$\mathbf{R}_{\boldsymbol{\epsilon}} = \mathbf{R}_{\boldsymbol{\theta}} - \mathbf{R}_{\boldsymbol{\theta}}\mathbf{H}^T(\mathbf{H}\mathbf{R}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{R_v})^{-1}\mathbf{H}\mathbf{R}_{\boldsymbol{\theta}} \tag{17}$$

where $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$. Let us denote the average transmit power at node k by $W_k^2 = P_{x_k} = [\mathbf{R}_{\boldsymbol{\theta}}]_{k,k} + \sigma_k^2$. Then, we can express the power optimization problem as follows:

$$\begin{array}{ll} \min & \sum_{k=1}^K W_k^2 \alpha_k \\ \text{s.t.} & [\mathbf{R}_{\boldsymbol{\epsilon}}]_{kk} \leq D_k,\ k = 1, \ldots, K. \end{array} \tag{18}$$

The optimization in (18) finds the power gain allocations that result in minimum total transmit power such that a maximum distortion level of D_k is allowed for node k. Using the matrix inversion lemma and after some simple algebraic operations, we can rewrite the optimization problem as

$$\begin{array}{ll} \min & \sum_{k=1}^K W_k^2 \alpha_k \\ \text{s.t.} & e_k^T\left[(\boldsymbol{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1}\right]e_k \leq D_k,\ k = 1, \ldots, K \end{array} \tag{19}$$

where $\boldsymbol{\Gamma} = \mathbf{H}^T\mathbf{R}_\mathbf{v}^{-1}\mathbf{H}$, and e_k is a $K \times 1$ vector whose k^{th} entry is one and all other entries are 0. Note that the distributed MMSE is efficient since the covariance matrix in (19) is equal to the Cramer-Rao bound given in (7).

The constraint in (19) defines a convex set. However, the optimization in (19) is not convex over α_k, $k = 1, \ldots, K$. By defining

$$r_k = \frac{\alpha_k g_k}{\alpha_k g_k \sigma_k^2 + \xi_k^2}$$

we obtain an equivalent convex optimization problem

$$\begin{array}{ll} \min & \sum_{k=1}^K \frac{W_k^2 \xi_k^2}{g_k}\left(\frac{r_k}{1 - r_k\sigma_k^2}\right) \\ \text{s.t.} & e_k^T(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1}e_k \leq D_k,\ 0 \leq r_k < \frac{1}{\sigma_k^2}, \end{array} \tag{20}$$

for $k = 1, \ldots, K$, where $\mathbf{R} = \mathrm{diag}\{r_1, \ldots, r_k\}$. Using $\frac{\partial X^{-1}}{\partial x_{rs}} = -X^{-1}E_{rs}X^{-1}$, $e_r e_s^T = E_{rs}$, and $e_r^T X e_s = [X]_{rs}$, we obtain the following KKT conditions:

$$W_k^2\frac{\xi_k^2}{g_k}\left(\frac{1}{1 - r_k\sigma_k^2}\right)^2 - \sum_{l=1}^K \lambda_l\left[(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1}\right]_{lk}^2 = 0 \tag{21}$$

$$e_k^T(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1}e_k = D_k, \tag{22}$$

for $0 \leq r_k \leq 1/\sigma_k^2$, and $k = 1, \ldots, K$. A closed form expression for this problem is not tractable, however, we can resort to numerical techniques to solve for $2K$ unknowns λ_k and r_k. A simple and straightforward solution exists for the case where local observations are independent, i.e., $\mathbf{R}_{\boldsymbol{\theta}}$ is diagonal. For a general correlation model, we present several numerical results in Section IV.

Example 1: *Independent Observations:* Let $\mathbf{R}_{\boldsymbol{\theta}} = \text{diag}(\chi_1^2, \ldots, \chi_K^2)$. Then we have

$$r_k^{\text{opt}} = \left(\frac{1}{D_k} - \frac{1}{\chi_k^2}\right)^+ \tag{23}$$

for $D_k \le \chi_k^2$, and

$$\alpha_k^{\text{opt}} = \frac{\xi_k^2 r_k^{\text{opt}}}{g_k(1 - r_k^{\text{opt}}\sigma_k^2)} \tag{24}$$

Note that because of the observation noise, D_k is lower bounded by $\frac{\sigma_k^2\chi_k^2}{\sigma_k^2+\chi_k^2}$.

2) Average Distortion Constraint: If one seeks to satisfy an average square error distortion across all sensor nodes, we can express the power optimization problem as follows:

$$\begin{array}{ll} \min & \sum_{k=1}^K W_k^2\alpha_k \\ \text{s.t.} & \frac{1}{K}\text{tr}(\mathbf{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \le D_0 \end{array} \tag{25}$$

where $\mathbf{\Gamma} = \mathbf{H}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{H}$. Note that the distributed MMSE is efficient since the covariance matrix in (25) is equal to the Cramer-Rao bound given in (7). Observing that $\text{tr}(X^{-1})$ is convex over the set of positive definite matrices $\{X : X > 0\}$, and that $\mathbf{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1} > 0$, e.g., the sum of two positive definite matrices is also a positive definite matrix, we conclude that the constraint in (25) defines a convex set. Once again, we define

$$r_k = \frac{\alpha_k g_k}{\alpha_k g_k \sigma_k^2 + \xi_k^2}$$

to convert the optimization in (25) to a convex one over r_k, $k = 1, \ldots, K$:

$$\begin{array}{ll} \min & \sum_{k=1}^K \frac{W_k^2\xi_k^2}{g_k}\left(\frac{r_k}{1-r_k\sigma_k^2}\right) \\ \text{s.t.} & \frac{1}{K}\text{tr}(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \le D_0,\ 0 \le r_k < \frac{1}{\sigma_k^2} \end{array} \tag{26}$$

where $\mathbf{R} = \text{diag}\{r_1, \ldots, r_k\}$. Now, we can use the Lagrangian method where we can obtain KKT conditions as:

$$W_k^2\frac{\xi_k^2}{g_k}\left(\frac{1}{1 - r_k\sigma_k^2}\right)^2 - \lambda_0\frac{1}{K}\left[(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-2}\right]_{kk} = 0 \tag{27}$$

$$\frac{1}{K}\text{tr}(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} = D_0 \tag{28}$$

for $0 \le r_k \le 1/\sigma_k^2$. Equations (27) and (28) define a set of $2K$ equations with $2K$ unknowns and one can solve for the unknowns with numerical techniques. If the local observations are independent, i.e., $\mathbf{R}_{\boldsymbol{\theta}}$ is diagonal, the optimization in (26) assumes a closed form solution as shown in the following example. For a general correlation model, we resort to numerical techniques as shown in Section IV.

Example 2: *Independent Observations:* Let $\mathbf{R}_{\boldsymbol{\theta}} = \text{diag}(\chi_1^2, \ldots, \chi_K^2)$, and without loss of generality, assume that $\frac{W_1\xi_1}{\sqrt{g_1}\chi_1^2} \le \cdots \le \frac{W_K\xi_K}{\sqrt{g_K}\chi_K^2}$. We define

$$A(J) = \frac{KD_0 - \text{tr}(\mathbf{R}_{\boldsymbol{\theta}}) + \sum_{j=1}^J \frac{\chi_j^4}{\chi_j^2+\sigma_j^2}}{\sum_{j=1}^J \frac{\chi_j^2 W_j\xi_j}{(\chi_j^2+\sigma_j^2)\sqrt{g_j}}} \tag{29}$$

and $f(J) = \frac{W_J\xi_J}{\sqrt{g_J}\chi_J^2}A(J)$, and then determine the unique J_1 such that $f(J_1) \le 1$ and $f(J_1+1) > 1$. Simplifying the KKT conditions in (27) and (28) and solving for r_j, we finally arrive at

$$r_j^{\text{opt}} = \begin{cases} \dfrac{1 - \frac{W_j\xi_j}{\sqrt{g_j}\chi_j^2}A(J_1)}{\sigma_j^2 + \frac{W_j\xi_j}{\sqrt{g_j}}A(J_1)} & j = 1, \ldots, J_1 \\ 0 & j = J_1+1, \ldots, K \end{cases} \tag{30}$$

Therefore, we have the optimal power gains

$$\alpha_j^{\text{opt}} = \frac{\xi_j^2}{g_j}\frac{r_j^{\text{opt}}}{1 - r_j^{\text{opt}}\sigma_j^2} \qquad \blacksquare \tag{31}$$

A power-distortion region, in a similar fashion to the the theoretically optimum power distortion region described in Section III-A, can also be defined for the uncoded transmission system with MMSE. For example, for the average distortion characterization, we have

$$\mathcal{P}^{\text{MMSE}}(D_0) = \left\{P_1, \ldots, P_K : D_0 \ge \frac{1}{K}\text{tr}(\mathbf{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1}\right\} \tag{32}$$

where $P_k = \alpha_k([\mathbf{R}_{\boldsymbol{\theta}}]_{kk} + \sigma_k^2)$. In Section IV, several comparisons between the power-distortion regions achieved by different schemes are also provided. We note that if the observations in different sensor nodes are uncorrelated Gaussian, it is easy to show that $\mathcal{P}^{\text{MMSE}}(D_0) = \mathcal{P}(D_0)$, i.e., uncoded transmission with MMSE at the receiver achieves the theoretically optimum performance.

C. Best Linear Unbiased Estimation and Optimal Power Allocation

Note that MMSE requires the knowledge of the autocorrelation matrix $\mathbf{R}_{\boldsymbol{\theta}}$. Therefore, MMSE can not be employed if we are not able to estimate the source correlation matrix or we do not have access to that information, that is, if we do not have any statistical knowledge on the source. The best linear unbiased estimator (BLUE) does not require the statistical knowledge of $\boldsymbol{\theta}$ and is given by

$$\hat{\boldsymbol{\theta}} = \left[\mathbf{H}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{y}.$$

The mean-squared error for this estimator can be obtained as

$$\mathbf{R}_{\boldsymbol{\epsilon}} = \left[\mathbf{H}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{H}\right]^{-1} \tag{33}$$

$$= \text{diag}\{\sigma_1^2 + \frac{\xi_1^2}{\alpha_1 g_1}, \ldots, \sigma_K^2 + \frac{\xi_K^2}{\alpha_K g_K}\} \tag{34}$$

We will again seek the optimal power allcoation for the two distortion criterion discussed above.

1) Individual Power Constraint:: The minimum-energy power allocation problem based on BLUE can be expressed as

$$\begin{array}{ll} \min & \sum_{k=1}^K W_k^2\alpha_k \\ \text{s.t.} & \sigma_k^2 + \frac{\xi_k^2}{g_k\alpha_k} \le D_k \end{array} \tag{35}$$

for $\alpha_k \ge 0$, $k = 1, \ldots, K$. By defining $r_k = \sigma_k^2 + \frac{\xi_k^2}{g_k\alpha_k}$, we can convert this optimization problem to a convex one over r_k:

$$\begin{array}{ll} \min & \sum_{k=1}^K W_k^2\frac{\xi_k^2}{g_k(r_k - \sigma_k^2)} \\ \text{s.t.} & r_k \le D_k,\ r_k \ge \sigma_k^2, k = 1, \ldots, K. \end{array} \tag{36}$$

for $r_k \geq 0$. Using Lagrangian method, we find the optimal power allocation coefficients as

$$\alpha_k = \frac{\xi_k^2}{g_k(D_k - \sigma_k^2)} \tag{37}$$

Note that D_k lies in the interval (σ_k^2, ∞) because of the observation noise.

2) Average Distortion Constraint: The minimum-energy power allocation problem for this case can be expressed as

$$\begin{array}{ll} \min & \sum_{k=1}^{K} W_k^2 \alpha_k \\ \text{s.t.} & \frac{1}{K}\sum_{k=1}^{K} \sigma_k^2 + \frac{\xi_k^2}{g_k\alpha_k} \leq D_0 \end{array} \tag{38}$$

for $\alpha_k \geq 0$, $k = 1, \ldots, K$. Using $r_k = \sigma_k^2 + \frac{\xi_k^2}{g_k\alpha_k}$ we have the equivalent problem

$$\begin{array}{ll} \min & \sum_{k=1}^{K} W_k^2 \frac{\xi_k^2}{g_k(r_k - \sigma_k^2)} \\ \text{s.t.} & \frac{1}{K}\sum_{k=1}^{K} r_k \leq D_0,\ r_k \geq \sigma_k^2, k = 1, \ldots, K \end{array} \tag{39}$$

for $r_k \geq 0$. Computing the Lagrangian and after some manipulations, we have

$$\alpha_k^{\text{opt}} = \frac{\sqrt{\frac{\xi_k^2}{W_k^2 g_k}} \sum_{k=1}^{K} \sqrt{\frac{W_k^2 \xi_k^2}{g_k}}}{K D_0 - \sum_{k=1}^{K} \sigma_k^2} \tag{40}$$

We note that the estimation problem described here is different from the one considered in [17]. In [17], the data observed at each sensor node is assumed to be exactly the same while here we consider spatially varying data. As a result, the optimized power allocation for these two problems has different solutions. In [17], the optimal power allocation might result in turning off some of the sensors, while in case of spatially varying data, each sensor has to transmit its observation with a power proportional to the inverse of square root of the channel SNR, where $SNR = W_k^2 g_k / \xi_k^2$.

IV. Numerical Examples

In this section, we present numerical examples for the $\mathcal{P} - \mathcal{D}$-region and power allocation problems studied in the previous sections. The power allocation problem for the distributed MMSE estimation does not allow for a closed form expression, so we resort to the numerical techniques to solve for optimal r_k and then find the optimal α_k, $k = 1, \ldots, K$. The optimization for the distributed BLUE has a closed from solution and is given by (37) or (40).

We consider the spatial correlation model defined by a correlation matrix

$$[\mathbf{R}_{\boldsymbol{\theta}}]_{i,j} = \rho^{|j-i|},\ \rho < 1. \tag{41}$$

This matrix has a symmetric tridiagonal inverse that can be computed by

$$[\mathbf{R}_{\boldsymbol{\theta}}^{-1}]_{i,j} = \begin{cases} \frac{1}{1-\rho^2} & i = j = 1,\ K \\ \frac{1+\rho^2}{1-\rho^2} & 2 \leq i = j \leq K-1 \\ \frac{-\rho}{1-\rho^2} & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases} \tag{42}$$

Fig. 1 depicts the power-distortion region that can be achieved by the optimum system and the uncoded system with MMSE and BLUE assuming an average distortion level of $D_0 = 0.2$ with $K = 2$ sensors when $\rho = 0.9$. From the figure, we observe that the MMSE performs significantly better than the BLUE estimator. It is also clear that $P(D_0) \supseteq P^{\text{MMSE}}(D_0) \supseteq P^{\text{BLUE}}(D_0)$. It is seen that we can satisfy an average distortion level of 0.2 when either $P_1 \geq 10$, or $P_2 \geq 8$, and this can be achieved by an uncoded transmission with MMSE at the receiver. This implies that explointing the correlation between the observations we can estimate one of them from the other one to attain a certain average distortion level. In Fig. 2, we depict power-distortion regions for the same system assuming a vector distortion level of $\mathbf{D} = [0.2\ 0.2]$. For this case, we also have $P(D_0) \supseteq P^{\text{MMSE}}(D_0) \supseteq P^{\text{BLUE}}(D_0)$. A remarkable observation is that even if the correlation coefficient between the sensor observations is as high as $\rho = 0.9$, and one seeks a certain distortion level for each estimation, each sensor node should be transmitting at a non-zero power level regardless of the power level of the other one.

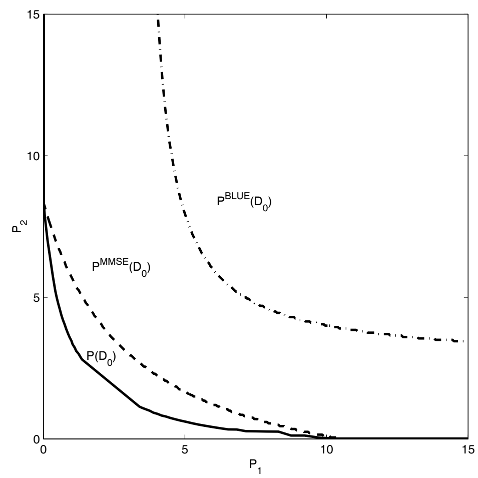

Fig. 1. Comparison of power-distortion regions attained by the optimum system and the uncoded transmission for the single-letter distortion case. Simulation parameters are $K = 2$, $\sigma_1^2 = \sigma_2^2 = 0.01$, $\mathbf{g} = [0.8,\ 1]$, $D_0 = 0.2$, $\rho = 0.9$.

Next, we provide a representative example for the optimum power allocation problem solved in Section III. We assume the following values for the parameters in our simulations: The observation noise variance, $\sigma_k^2 = 0.01$, and the channel noise variance, $\xi_k^2 = 1$ for $\forall k$. The correlation matrix of the observations is given by (41), e.g., $\chi_k^2 = 1$, $k = 1, \ldots, K$. In Figure 3, for a network with 2 sensor nodes, we illustrate the power vs. distortion performance for three cases: (i) The theoretically possible minimum power level obtained by (15) to satisfy the average distortion level of $D_0 = 0.2$, (ii) the minimum power required with MMSE, and (ii) the minimum power required with BLUE. The optimal power gains are obtained by solving (27) and (28). It is seen that, to satisfy some mean-squared error level D_0, the MMSE requires less transmission power than the BLUE does at all distortion levels. We also observe that at sufficiently large transmission power (to achieve distorion levels less than 0.02), the amplify-and-forward approach attains the distortion level of the optimum scheme (the difference becomes negligible). From the curves for $\rho = 0.5$ and $\rho = 0.9$, there is some gap between the

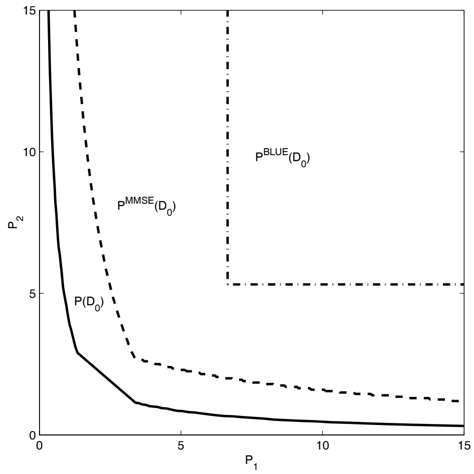

Fig. 2. Comparison of power-distortion regions attained by the optimum system and the uncoded transmission for the multiple-letter distortion case. Simulation parameters are $K = 2$, $\sigma_1^2 = \sigma_2^2 = 0.01$, $\mathbf{g} = [0.8,\ 1]$, $\mathbf{D} = [0.2\ 0.2]$, $\rho = 0.9$.

performance of the optimum uncoded scheme and that of a theoretically optimal scheme, and this performance gap is small for lower ρ values. As mentioned earlier, if $\rho = 0$, e.g., when the observations are uncorrelated, this gap disappears and the uncoded transmission achieves best performance.

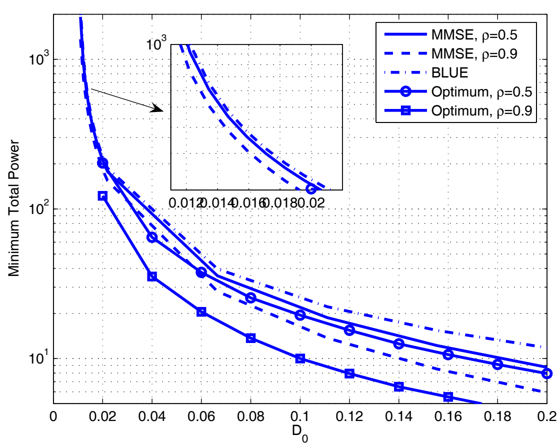

Fig. 3. D_0 vs. P for distributed estimation of of a correlated field with optimal power allocation with MMSE and BLUE. Simulation parameters: $\sigma_k^2 = 0.01$, $\xi_k^2 = 1$, $\rho = 0.5$ and 0.9, $K = 2$, $\mathbf{g} = [0.81]$

V. Conclusions

In this paper, we addressed the energy-efficient distributed estimation problem for correlated data in sensor networks. We derived the achievable power-distortion region for the distributed estimation problem where a fusion center tries to estimate a distributed field by a network of sensor nodes distributed in a terrain. We considered two different distortion characterization (a single- and a multiple-letter distortion measures) and and studied an uncoded analog transmission scheme where the noise-corrupted sensor observations are simply amplified and forwarded to the fusion center through the noisy flat-fading channel. Two different estimation techniques are investigated for the fusion center: (i) the minimum mean-square error estimation, which requires the knowledge of the correlation matrix, and (ii) minimum linear unbiased estimation, which does not require/exploit any statistical knowledge. For both estimation techniques, we determined the optimal power allocation scheme and the minimum required power with which one can satisfy a certain mean-squared error distortion level. Performance comparisons of the various schemes indicate that one needs to exploit the intersensor correlations for better energy efficiency. Comparison with optimal schemes indicate that as correlations between the sensor observation becomes small, the performance gap between the uncoded scheme and the theoretically optimum scheme decreases. Furthermore, comparions between the MMSE and BLUE performance indicates that exploiting the correlation among sensor observations (by MMSE) reduces the required power to attain some distortion level.

References

[1] Y.Sankarasubramaniam I.F.Akyildiz, W.Su and E.Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, pp. 102–114, August 2002.

[2] C. E. Shannon, "A mathematical thoery of communications," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, October 1948.

[3] T. Berger, *Rate Distortion Theory*, Prentice Hall, Englewood Cliffs, NJ, 1971.

[4] D. Slepian anf J. Wolf, "Noisless coding of correlated sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, Jan 1976.

[6] T. Berger and R. Yeung, "Multirerminal source coding with one distortion criterion," *IEEE Transactions on Information Theory*, vol. 35, pp. 228–236, Mar 1989.

[7] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Transactions on Information Theory*, vol. 45, pp. 106–117, Jan 1999.

[8] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 887–902, May 1996.

[9] H. Viswanathan and T. Berger, "The quadratic gaussian ceo problem," *IEEE Transactions on Information Theory*, vol. 43, no. 9, pp. 1549–1559, Sep 1997.

[10] Y. Oohama, "The rate-distortion function for the quadratic ceo problem," *IEEE Transactions on Information Theory*, vol. 44, pp. 1057–1070, May 1998.

[11] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic gaussian ceo problem," *Proceedings of the International Symposium on Information Theory*, p. 199, June 2004.

[12] J.-J. Xiao and Z.-Q. Luo, "Multiterminal source-channl communication under orthogonal multiple access channels," submitted for publication, Mar 2005.

[13] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, no. 4, pp. 558–567, Oct 1965.

[14] R. G. Gallager, *Information theory and realiable communication*, John Wiley and Sons, New York, NY, 1968.

[15] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[16] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," *Lecture Notes in Computer Science*, vol. 2634, pp. 162–177, April 2003.

[17] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Energy-efficient joint estimation in sensor networks: Analog vs. digital," in *ICASSP*, Philadelphia, PA, March 2005, IEEE, pp. 745–748.

[18] A. Jindal and K. Psounis, "Modeling spatially-correlated sensor network data," in *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, IEEE SECON 2004*, 2004, pp. 162–171.

[19] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Wiely, New York, 1968.

[20] J. Chen, Z. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion functions and its corresponding rate allocation schemes for the ceo problem," *IEEE Journal on Selected Areas in Comm.*, vol. 22, no. 6, pp. 977–987, Aug 2004.