Power Constrained Distributed Estimation With Correlated Sensor Data

Jun Fang and Hongbin Li

***Abstract*—We consider distributed estimation of a random scalar parameter in a power constrained wireless sensor network (WSN), where the measurements are sent from the sensors to the fusion center (FC) over noisy wireless channels by employing an analog transmission scheme. We study the power allocation problem with generally correlated sensor observations that can accommodate nonlinear measurement models and spatially correlated observation noise. An effective solution is developed by utilizing a tractable lower bound of the objective function. The proposed algorithm is also extended for random field estimation. Simulation results are presented to illustrate the effectiveness of the proposed algorithm.**

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

I. Introduction

Distributed parameter estimation is one of the fundamental problems arising from the wide applications of WSNs, in which a random variable or a random field is observed by multiple sensors whose observations are processed and transmitted to a fusion center (FC) to form an estimate of the parameter(s). This problem has attracted much attention over the past few years. As sensors in a network are powered by small-size batteries whose energy resource is severely limited, energy constraint is a primary issue that need to be taken into account in designing distributed estimation algorithms. A multitude of studies along this line have appeared recently, e.g., distributed estimation by using aggressive quantization strategies [1]–[3] or by employing a linear compression matrix [4]–[6] to reduce the transmission requirement. In most of the above studies, ideal wireless channels are assumed through which the measurements are sent from the sensors to the FC without distortion. This assumption, however, is undermined in practice due to the link noise and adverse channel effects such as fading. Increasing transmission power is one way to counteract channel impairments and improve the quality of the received signal. However, as the energy resources provided by the sensor networks are extremely limited, a power allocation problem naturally arises. The objective of power allocation is to find an optimum strategy to assign power among different sensors, aiming at minimizing the estimation error under certain transmission power constraints or its converse: satisfying a target distortion performance with a minimum energy consumption. Such a problem has been extensively investigated in [7]–[11]. In particular, the power scheduling was studied in [7] and [8] under the frameworks of an uncoded quadrature amplitude modulated (QAM) transmission strategy and an analog transmission strategy, respectively. The extension of [8] to distributed estimation of a random field was considered in [9]. In [10], the authors examined the power allocation in a multihop scenario where the information is forwarded from sensors to an FC via a routing tree. Recently, the minimal energy distributed estimation was discussed in [11], where the observation noise variance was assumed unknown and described by a statistical model. All these works, however, assume independent observation noise across the sensors. Also, although distributed estimation of random signals can be dealt with by [8] and [9], only linear signal-plus-noise models were considered.

In this correspondence, we examine the power allocation with arbitrarily spatially correlated sensor observations. Correlation among sensor observations could arise as a result of colored observation noise, which is usually the case when sensors are densely deployed in a finite area and the phenomena to be measured is subject to similar disturbance and correlated ambient noise. Moreover, even with independent observation noise, sensor observations are correlated when the signal to be estimated is modeled as a random variable or dependent random variables. In this case, most existing methods, e.g., [8] and [9], address only linear signal-plus-noise observation models. This correspondence considers a more general case with correlated observations resulting from either linear or nonlinear observation models. A main difficulty arising from generally correlated sensor observations for optimal power allocation is that the underlying optimization problem involves a nondiagonal matrix inverse that is usually hard to deal with. To circumvent this difficulty, we utilize the Cauchy–Schwarz inequality to find a tractable lower bound of the objective function to be maximized. A suboptimal solution is sought for by maximizing the derived lower bound. The new optimization turns out to be a convex quadratic programming (QP) problem, which can be efficiently solved.

II. Problem Formulation

A. Random Scalar Parameter Estimation

Consider a WSN consisting of N spatially distributed sensors linked with a FC. Each sensor, say sensor n, makes an observation, x_n, that is correlated with the random signal of interest θ. The measurement model used to describe the underlying physical observation mechanism can take any form and is not confined to a standard linear signal-plus-noise model $x_n = h_n\theta + w_n$ (here, h_n is a scaling factor). The measurement model, for example, can be characterized by $x_n = f_n(\theta) + w_n \; n = 1, 2, \ldots, N$, where $\{f_n(\cdot)\}$ are general nonlinear functions representing the relationship between the signal and the measurements, $\{w_n\}$ is the observation noise that could be independent or spatially correlated. We assume that the signal θ and the data $\{x_n\}$ have zero mean with autocorrelation and cross-correlation σ_θ^2, $\mathbf{R}_{\theta x}$, and $\mathbf{R}_x$ available at the FC, where we define $\mathbf{R}_{\theta x} \triangleq E[\theta \mathbf{x}^T]$, $\mathbf{R}_x \triangleq E[\mathbf{x}\mathbf{x}^T]$, in which $\mathbf{x} \triangleq [x_1 \; x_2 \; \ldots \; x_N]^T$. Note that the zero-mean assumption, which is made for simplicity for exposition, does not sacrifice the generality. *A priori* knowledge of the auto- and cross-correlation can come either from specific data models or from sample estimation after a training phase [6].

We adopt a simple uncoded analog amplify-and-forward scheme [8] to transmit the observations $\{x_n\}$ to a FC. The observations are transmitted to the FC via independent additive white Gaussian noise (AWGN) channels that can be realized by a multiaccess technique such as TDMA or FDMA. The received signal from the n^{th} channel (or sensor n), therefore, is given by

$$y_n = g_n \alpha_n x_n + v_n \qquad n = 1, 2, \ldots, N \tag{1}$$

where α_n is the amplification factor employed at sensor n, g_n and v_n denote the channel gain and the additive channel noise associated with sensor n, respectively. The channel noise $\{v_n\}$ is assumed statistically

Manuscript received July 25, 2008; accepted February 26, 2009. First published March 31, 2009; current version published July 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mats Bengtsson. This work was supported in part by the National Science Foundation under Grant CCF-0514938 and by the Air Force Office of Scientific Research under Grant FA9550-09-1-0310.

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

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

Digital Object Identifier 10.1109/TSP.2009.2020033

independent of the sensor observations $\{x_n\}$, with zero mean and covariance matrix $\mathbf{R}_v \triangleq E[\mathbf{v}\mathbf{v}^T] = \mathrm{diag}(\sigma_{v,1}^2, \sigma_{v,2}^2, \ldots, \sigma_{v,N}^2)$, where $\mathbf{v} \triangleq [v_1\ v_2\ \ldots\ v_N]^T$, and $\sigma_{v,n}^2$ denotes the variance of the channel noise v_n.

Let $\mathbf{y} \triangleq [y_1\ y_2\ \ldots\ y_N]^T$, we rewrite (1) in a vector form:

$$\mathbf{y} = \mathbf{D}_h \mathbf{x} + \mathbf{v} \tag{2}$$

where $\mathbf{D}_h \triangleq \mathrm{diag}(g_1\alpha_1, g_2\alpha_2, \ldots, g_N\alpha_N)$. Using the received data $\mathbf{y}$ at the FC, the linear minimum mean-square error (LMMSE) estimate of θ is known as [12]

$$\hat{\theta} = \mathbf{R}_{\theta y}\mathbf{R}_y^{-1}\mathbf{y} = \mathbf{R}_{\theta x}\mathbf{D}_h^T \left(\mathbf{D}_h \mathbf{R}_x \mathbf{D}_h^T + \mathbf{R}_v\right)^{-1} \mathbf{y} \tag{3}$$

and the estimation mean-square error (MSE) is given by

$$\begin{aligned} E\left[(\theta - \hat{\theta})^2\right] &= \sigma_\theta^2 - \mathbf{R}_{\theta y}\mathbf{R}_y^{-1}\mathbf{R}_{\theta y}^T \\ &= \sigma_\theta^2 - \mathbf{R}_{\theta x}\mathbf{D}_h^T \left(\mathbf{D}_h \mathbf{R}_x \mathbf{D}_h^T + \mathbf{R}_v\right)^{-1} \mathbf{D}_h \mathbf{R}_{\theta x}^T \\ &= \sigma_\theta^2 - \mathbf{R}_{\theta x}\mathbf{D}_\alpha^T \left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right)^{-1} \mathbf{D}_\alpha \mathbf{R}_{\theta x}^T \end{aligned} \tag{4}$$

where $\mathbf{R}_{\theta y} \triangleq E[\theta \mathbf{y}^T]$, $\mathbf{R}_y \triangleq E[\mathbf{y}\mathbf{y}^T]$, $\mathbf{D}_\alpha \triangleq \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_N)$, and $\tilde{\mathbf{R}}_v \triangleq \mathrm{diag}((\sigma_{v,1}^2/g_1^2), (\sigma_{v,2}^2/g_2^2), \ldots, (\sigma_{v,N}^2/g_N^2))$ with its diagonal entries the inverse of the channel signal-to-noise ratios (SNRs).

We see that the LMMSE estimator involves determining a set of amplification factors $\{\alpha_n\}$ that are used to adjust the average transmission power of the sensors. Specifically, the average power of sensor n is given by

$$P_n = \alpha_n^2 \sigma_{x,n}^2, \qquad n = 1, 2, \ldots, N \tag{5}$$

where $\sigma_{x,n}^2$ denotes the variance of the sensor observation x_n. Due to the severely limited energy resources, these factors cannot be chosen arbitrarily large. Naturally, with the total transmission power constrained, we may wish to find an optimal set $\{\alpha_n\}$ to minimize the estimation MSE. By noting that

$$\sum_{n=1}^{N} P_n = \mathrm{tr}\left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T\right) \tag{6}$$

this power allocation problem is formulated as

$$\begin{aligned} \max_{\mathbf{D}_\alpha} \quad & \mathrm{tr}\left(\mathbf{R}_{\theta x}\mathbf{D}_\alpha^T \left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right)^{-1} \mathbf{D}_\alpha \mathbf{R}_{\theta x}^T\right) \\ \text{s.t.} \quad & \mathrm{tr}\left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T\right) = P_{\mathrm{total}} \end{aligned} \tag{7}$$

where the power constraint is represented as an equality instead of an inequality because it can be easily verified that the objective function is a monotonically increasing function of P_{total}. We assume the FC has the knowledge of the channel state information, $\{g_n\}$, and the channel noise variances. In practice, these information can be estimated from the training data from the sensors (e.g., [8]).

B. Random Field Estimation

The above formulation can be easily extended to the case where the random signal of interest, instead of a common scalar parameter, is a set of dependent, spatially distributed random variables, i.e., $\boldsymbol{\theta} \triangleq [\theta_1\ \theta_2\ \ldots\ \theta_N]^T$, where θ_n is an unknown parameter associated with sensor n. Again, the mapping between the observation x_n and the unknown θ_n is not restricted to a linear signal-plus-noise model: $x_n = h_n\theta_n + w_n$. It can be easily verified that the LMMSE estimator and the corresponding estimation covariance matrix have the same expressions of (3) and (4), with σ_θ^2 replaced by $\mathbf{R}_\theta \triangleq E[\boldsymbol{\theta}\boldsymbol{\theta}^T]$ and $\mathbf{R}_{\theta x} \triangleq E[\boldsymbol{\theta}\mathbf{x}^T]$. The formulation of the power allocation problem, therefore, remains the same as (7).

III. Proposed Approach

Power allocation problem was also studied in [8], [9]. However, they only addressed linear signal-plus-noise measurement models with independent observation noise. Specifically, the formulation of the estimation MSE or the estimation covariance matrix, based on which the methods [8], [9] are developed, are specific for linear models; also, only under the assumption of independent observation noise across the sensors, the estimation MSE can be simplified as an analytical form [8] or cast into a convex function [8], [9] (see Section IV for a detailed discussion). In contrast to [8] and [9], we study a general power allocation problem (7) that accommodates nonlinear measurement models and spatially correlated observation noise.

A main difficulty associated with (7) is that the underlying optimization problem involves a nondiagonal matrix inverse that is usually hard to deal with. To circumvent this difficulty, we, instead, propose to maximize a lower bound of the objective function in (7). We introduce the following lemma that provides a natural framework allowing us to find a lower bound of the above objective function.

Lemma 1: For any $\mathbf{G} \in \mathbb{R}^{q\times p}$ and positive-definite matrix $\mathbf{Q} \in \mathbb{R}^{q\times q}$, the following inequality holds:

$$\mathrm{tr}(\mathbf{G}^T\mathbf{Q}^{-1}\mathbf{G}) \geq \frac{\left(\mathrm{tr}(\mathbf{G}^T\mathbf{G})\right)^2}{\mathrm{tr}(\mathbf{G}^T\mathbf{Q}\mathbf{G})}. \tag{8}$$

Proof: Let $\mathbf{A} = \mathbf{G}^T\mathbf{Q}^{-(1/2)}$ and $\mathbf{B} = \mathbf{G}^T\mathbf{Q}^{1/2}$, Lemma 1 follows from the Cauchy–Schwarz inequality:

$$\mathrm{tr}(\mathbf{A}\mathbf{A}^T)\mathrm{tr}(\mathbf{B}\mathbf{B}^T) \geq \left(\mathrm{tr}(\mathbf{A}\mathbf{B}^T)\right)^2.$$

■

By utilizing Lemma 1, a lower bound of the cost function in (7) is provided below.

Proposition 1: For any arbitrary correlation matrix $\mathbf{R}_x$, we have

$$\begin{aligned} &\mathrm{tr}\left(\mathbf{R}_{\theta x}\mathbf{D}_\alpha^T \left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right)^{-1} \mathbf{D}_\alpha \mathbf{R}_{\theta x}^T\right) \\ &\qquad \geq \frac{\left(\mathrm{tr}\left(\mathbf{R}_{\theta x}\mathbf{D}_\alpha^T\mathbf{D}_\alpha\mathbf{R}_{\theta x}^T\right)\right)^2}{\mathrm{tr}\left(\mathbf{R}_{\theta x}\mathbf{D}_\alpha^T \left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right) \mathbf{D}_\alpha \mathbf{R}_{\theta x}^T\right)}. \end{aligned} \tag{9}$$

Proof: The result comes directly from Lemma 1 by letting $\mathbf{G} = \mathbf{D}_\alpha\mathbf{R}_{\theta x}^T$ and $\mathbf{Q} = \mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v$. ■

In the following, for simplicity, we first present our solution for the scalar parameter case and then extend it to a set of dependent random variables.

A. Scalar Parameter

The cross-correlation $\mathbf{R}_{\theta x}$ is a row vector when the signal of interest, θ, is a scalar random parameter. In this case, we can rewrite the lower bound and the power constraint into more tractable forms by exploiting the diagonal structure of $\mathbf{D}_\alpha$. Define

$$\mathbf{f} \triangleq \left[\alpha_1^2 \quad \alpha_2^2 \quad \ldots \quad \alpha_N^2\right]^T \tag{10}$$

it is easy to verify that

$$\mathbf{R}_{\theta x}\mathbf{D}_\alpha^T\mathbf{D}_\alpha\mathbf{R}_{\theta x}^T = \boldsymbol{\phi}_{\theta x}^T\mathbf{f} \tag{11}$$

$$\mathbf{R}_{\theta x}\mathbf{D}_\alpha^T \left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right) \mathbf{D}_\alpha \mathbf{R}_{\theta x}^T = \mathbf{f}^T\mathbf{P}\mathbf{f} + \boldsymbol{\phi}_v^T\mathbf{f} \tag{12}$$

$$\mathrm{tr}\left(\mathbf{D}_\alpha \mathbf{R}_x \mathbf{D}_\alpha^T\right) = \boldsymbol{\phi}_x^T\mathbf{f} \tag{13}$$

where $\boldsymbol{\phi}_{\theta x} \triangleq (\mathbf{R}_{\theta x} \odot \mathbf{R}_{\theta x})^T$, $\mathbf{P} \triangleq \mathrm{DIAG}(\mathbf{R}_{\theta x})\mathbf{R}_x\mathrm{DIAG}(\mathbf{R}_{\theta x})$, $\boldsymbol{\phi}_v \triangleq \mathbf{R}_{\theta x}^T \odot \mathrm{diag}(\tilde{\mathbf{R}}_v) \odot \mathbf{R}_{\theta x}^T$, and $\boldsymbol{\phi}_x \triangleq \mathrm{diag}(\mathbf{R}_x)$, in which $\odot$ denotes the Hadamard product, also known as the entrywise product, $\mathrm{DIAG}(\mathbf{a})$ is a diagonal matrix with its diagonal entries given by $\mathbf{a}$, $\mathrm{diag}(\mathbf{A})$ is a column vector with its entries formed by the diagonals of $\mathbf{A}$.

Using (9) and (11)–(13), we now turn to the following problem:

$$\begin{aligned} \max_{\mathbf{f}} \quad & \frac{\boldsymbol{\phi}_{\theta x}^T \mathbf{f}\mathbf{f}^T \boldsymbol{\phi}_{\theta x}}{\mathbf{f}^T\mathbf{P}\mathbf{f} + \boldsymbol{\phi}_v^T\mathbf{f}} \\ \text{s.t.} \quad & \boldsymbol{\phi}_x^T\mathbf{f} = P_{\text{total}} \\ & \mathbf{f} \geq \mathbf{0} \end{aligned} \tag{14}$$

where the inequality, $\mathbf{f} \geq \mathbf{0}$, comes from the fact that every entry of the vector $\mathbf{f}$ is greater than or equal to zero [see (10)].

We first consider a special case where $\boldsymbol{\phi}_x = c\boldsymbol{\phi}_{\theta x}$, in which c is a constant. This particular case arises from a linear model $x_i = \theta + w_i$ when the observation noise variances across the sensors are identical. In this case, the power constraint is equivalent to imposing a constant constraint: $\boldsymbol{\phi}_{\theta x}^T\mathbf{f} = P_{\text{total}}/c$ on the numerator of the objective function in (14). Hence, to maximize the objective function, we only need to minimize its denominator. Therefore, we reach the following quadratic programming (QP) problem:

$$\begin{aligned} \min_{\mathbf{f}} \quad & \mathbf{f}^T\mathbf{P}\mathbf{f} + \boldsymbol{\phi}_v^T\mathbf{f} \\ \text{s.t.} \quad & \boldsymbol{\phi}_x^T\mathbf{f} = P_{\text{total}} \\ & \mathbf{f} \geq \mathbf{0}. \end{aligned} \tag{15}$$

As $\mathbf{P}$ is positive definite, the QP problem (15) is convex and admits an unique global solution. Although the solution can hardly be found analytically, it can be efficiently solved by numerical methods such as the interior point method.

For the general case where $\boldsymbol{\phi}_x \neq c\boldsymbol{\phi}_{\theta x}$, to solve (14), we, firstly, consider the optimization of $\mathbf{f}$ by fixing the numerator of the objective function. Let $\lambda \triangleq \boldsymbol{\phi}_{\theta x}^T\mathbf{f}$. Clearly, for any specified $\lambda \in (\lambda_{\min}, \lambda_{\max})$, where $(\lambda_{\min}, \lambda_{\max})$ is a feasible region of λ, the optimum $\mathbf{f}$ is given by the following QP problem:

$$\begin{aligned} \min_{\mathbf{f}} \quad & \mathbf{f}^T\mathbf{P}\mathbf{f} + \boldsymbol{\phi}_v^T\mathbf{f} \\ \text{s.t.} \quad & \boldsymbol{\phi}_x^T\mathbf{f} = P_{\text{total}} \\ & \boldsymbol{\phi}_{\theta x}^T\mathbf{f} = \lambda \\ & \mathbf{f} \geq \mathbf{0}. \end{aligned} \tag{16}$$

The optimization (16), as (15), is convex and can be efficiently solved by numerical methods. Let $\mathbf{f}_{\text{opt}}(\lambda)$ denote the optimum solution of $\mathbf{f}$ associated with each λ. By substituting it into (17), we obtain the following one-dimensional search of λ:

$$\begin{aligned} \max_{\lambda} \quad & \frac{\lambda^2}{\mathbf{f}_{\text{opt}}^T(\lambda)\mathbf{P}\mathbf{f}_{\text{opt}}(\lambda) + \boldsymbol{\phi}_v^T\mathbf{f}_{\text{opt}}(\lambda)} \\ \text{s.t.} \quad & \lambda_{\max} > \lambda > \lambda_{\min}. \end{aligned} \tag{17}$$

To find an optimum λ, we choose a set of candidates $\{\lambda_k\}_{k=1}^M$ from $(\lambda_{\min}, \lambda_{\max})$. The best λ in the set $\{\lambda_k\}_{k=1}^M$ is determined as the one that maximizes the cost function of (17). Therefore, the solution to (14) is finally given by $\{\lambda^*, \mathbf{f}_{\text{opt}}(\lambda^*)\}$, where λ^* denotes the best λ chosen from the candidate set.

We note that a sum power constraint imposes a constraint on power consumption at the network level and therefore, has a direct impact on the network lifespan. Meanwhile, in some application scenarios, a per-node based power constraint may also be desirable in order to prevent excessive battery drainage for individual sensors. Our proposed method can be readily extended to incorporate individual power constraints for all sensors. Recalling (5), the individual power constraints can be expressed by the inequality: $\boldsymbol{\eta} \geq \mathbf{f}$, where $\boldsymbol{\eta} \triangleq [\eta_1\ \eta_2\ \ldots\ \eta_N]^T$, $\eta_n \triangleq P_{\max,n}/\sigma_{x,n}^2$, and $P_{\max,n}$ denotes the maximum power consumption for sensor n. By incorporating this constraint, the optimization (14) becomes

$$\begin{aligned} \max_{\mathbf{f}} \quad & \frac{\boldsymbol{\phi}_{\theta x}^T \mathbf{f}\mathbf{f}^T \boldsymbol{\phi}_{\theta x}}{\mathbf{f}^T\mathbf{P}\mathbf{f} + \boldsymbol{\phi}_v^T\mathbf{f}} \\ \text{s.t.} \quad & \boldsymbol{\phi}_x^T\mathbf{f} = P_{\text{total}} \\ & \boldsymbol{\eta} \geq \mathbf{f} \geq \mathbf{0} \end{aligned} \tag{18}$$

which can be solved by following the same approach used to solve (14).

B. Random Field

Consider the random signal of interest is a set of dependent, spatially distributed random variables: $\boldsymbol{\theta} \triangleq [\theta_1\ \theta_2\ \ldots\ \theta_N]^T$. In this case, the cross-correlation $\mathbf{R}_{\theta x}$ is an $N \times N$ matrix instead of a row vector. Nevertheless, by utilizing

$$\mathrm{tr}\left(\mathbf{R}_{\theta x}\mathbf{A}\mathbf{R}_{\theta x}^T\right) = \sum_{n=1}^{N} \mathbf{R}_{\theta x}[n,:]\mathbf{A}\mathbf{R}_{\theta x}[n,:]^T \tag{19}$$

and following the derivation of (11) and (12) for each component $\mathbf{R}_{\theta x}[n,:]\mathbf{A}\mathbf{R}_{\theta x}[n,:]^T$, the lower bound given in (9) can be turned into a same form as the objective function given in (14), where $\mathbf{R}_{\theta x}[n,:]$ denotes the nth row of $\mathbf{R}_{\theta x}$. Therefore, we encounter the same problem as (14) which can be solved by following the same way as described in previous subsection.

IV. Discussions

To better understand the difference between our work and [8] and [9], we explain why the methods [8], [9] are only for linear measurement models. Specifically, we focus our discussion on [8] since the extension of our discussion to [9] is straightforward. We note that the method [8], when θ is modeled as a random parameter, relies on the following reformulation of the estimation MSE (4) which is based on the linear relationship $\mathbf{x} = \mathbf{h}\theta + \mathbf{w}$:

$$\begin{aligned} & E\left[(\theta - \hat{\theta})^2\right] \\ &= \sigma_\theta^2 - \mathbf{R}_{\theta x}\mathbf{D}_\alpha^T\left(\mathbf{D}_\alpha\mathbf{R}_x\mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right)^{-1}\mathbf{D}_\alpha\mathbf{R}_{\theta x}^T \\ &= \sigma_\theta^2 - \sigma_\theta^4\mathbf{h}^T\mathbf{D}_\alpha^T \cdot \left(\mathbf{D}_\alpha\left(\sigma_\theta^2\mathbf{h}\mathbf{h}^T + \mathbf{R}_w\right)\mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right)^{-1}\mathbf{D}_\alpha\mathbf{h} \\ &= \left(\sigma_\theta^{-2} + \mathbf{h}^T\mathbf{D}_\alpha^T\left(\mathbf{D}_\alpha\mathbf{R}_w\mathbf{D}_\alpha^T + \tilde{\mathbf{R}}_v\right)^{-1}\mathbf{D}_\alpha\mathbf{h}\right)^{-1}. \end{aligned} \tag{20}$$

If the observation noise is independent, i.e., $\mathbf{R}_w$ is diagonal, the above reformulation circumvents a direct nondiagonal matrix inversion and yields an analytical solution. However, for generally nonlinear measurement models, the above reformulation does not lead to a simplification because the matrices $\mathbf{R}_x$ and $\mathbf{R}_{\theta x}$ usually do not satisfy a relationship that holds valid for linear models with independent observation noise, i.e., $\mathbf{R}_x = c\mathbf{R}_{\theta x}^T\mathbf{R}_{\theta x} + \boldsymbol{\Lambda}$, where c is a positive constant and $\boldsymbol{\Lambda}$ is a diagonal matrix with nonnegative diagonal elements. Hence, the proposed method [8] is not applicable to generally nonlinear models.

Nonlinear model occurs in many sensor network applications. For example, in the target tracking scenario, the observation equations are nonlinear because there is a natural power law decay of the sensed information with distance from the target. In addition, the physical sensing mechanism can be designed to be a nonlinear function to compress the signal amplitude in order to capture the characteristic that small signal amplitudes occur more frequently than large ones. One

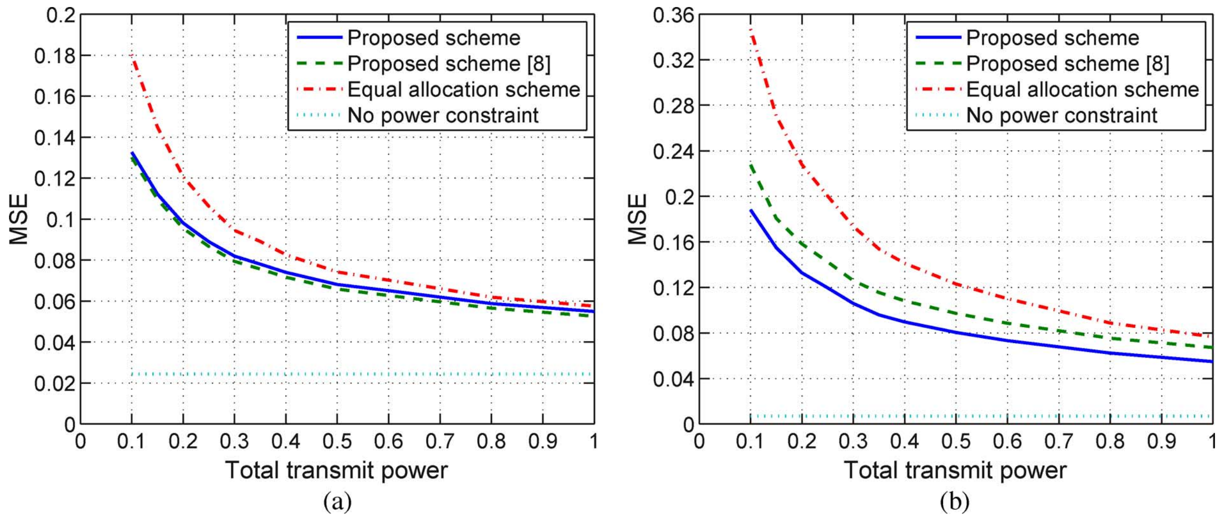

Fig. 1. Example A: MSEs versus total transmit power for a linear model with uncorrelated and correlated observation noise, respectively. (a) $\beta = 0$ and (b) $\beta = -0.9$.

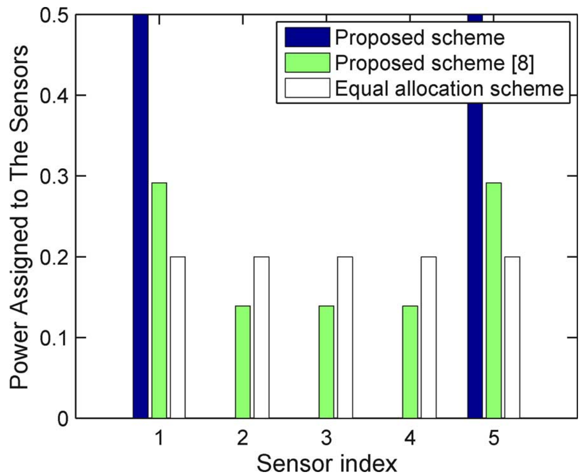

Fig. 2. Example A: A bar graph depicting the power allocation of the three schemes.

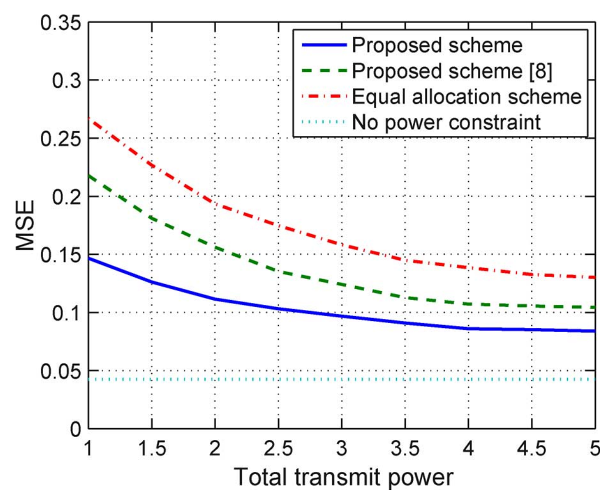

Fig. 3. Example A: MSEs versus total transmit power for a nonlinear model.

may argue that the sensor can carry out a local function reverse to remove the nonlinearity. However, this imposes a high computational demand to the sensors. Also, this is only feasible when the local nonlinear mapping from the signal to the observation is unique, i.e., each possible signal value corresponds to a unique observation value, which may not be true in practice.

V. Simulation Results

We present simulation results to illustrate the estimation performance of the proposed algorithm. Two subsections are included, in which the estimation of a random scalar parameter and a random field are considered, respectively.

A. Estimation of a Random Parameter

We first examine the following linear signal-plus-noise model: $x_n = \theta + w_n$, where the observation noise $\{w_n\}$ is modeled as an autoregressive (AR)(1) model, which is given by

$$w_n = \beta w_{n-1} + u_n \qquad n = 2, \ldots, N \tag{21}$$

in which β is the AR coefficient; $\{u_n\}$ is a white noise process with zero mean and variance $\sigma_u^2 = 0.5$. We compare our method with the proposed power allocation method [8] which is implemented by ignoring the off-diagonal elements of the noise correlation matrix, and an equal power allocation scheme where all sensors transmit the same amount of power. In our simulations, we set $N = 20$, and $\sigma_\theta^2 = 1$. The channel SNR is given by

$$\mathrm{SNR}_{\mathrm{ch},n} = 10 \log_{10} \frac{g_n^2}{\sigma_{v,n}^2}, \quad n = 1, \ldots, N \tag{22}$$

where $\{g_n\}$ are i.i.d. normal random variables with zero mean and unit variance, and we choose $\sigma_{v,n}^2 = 0.01$ for any n. Fig. 1 shows the estimation MSEs of the three schemes as a function of the total transmit power P_{total}, where the AR coefficient β in (21) is set to be 0 and -0.9 respectively, with $\beta = 0$ corresponding to an independent scenario and $\beta = -0.9$ corresponding to a correlated scenario. The results are averaged over 500 Monte Carlo runs, with the channel gains $\{g_n\}$ randomly generated for each run. The minimum MSE achieved without imposing any power constraint, i.e., by assuming ideal wireless links, is also included. We note that, when $\beta = 0$, the method [8] is optimal and yields the best performance among the three schemes. Nevertheless, our proposed method achieves quasi-optimal performance with only a slight loss as compared with [8]. When $\beta = -0.9$, [8] is no longer optimal. In this case, our proposed method presents a performance advantage over the method [8] and the equal power allocation scheme. To meet the same

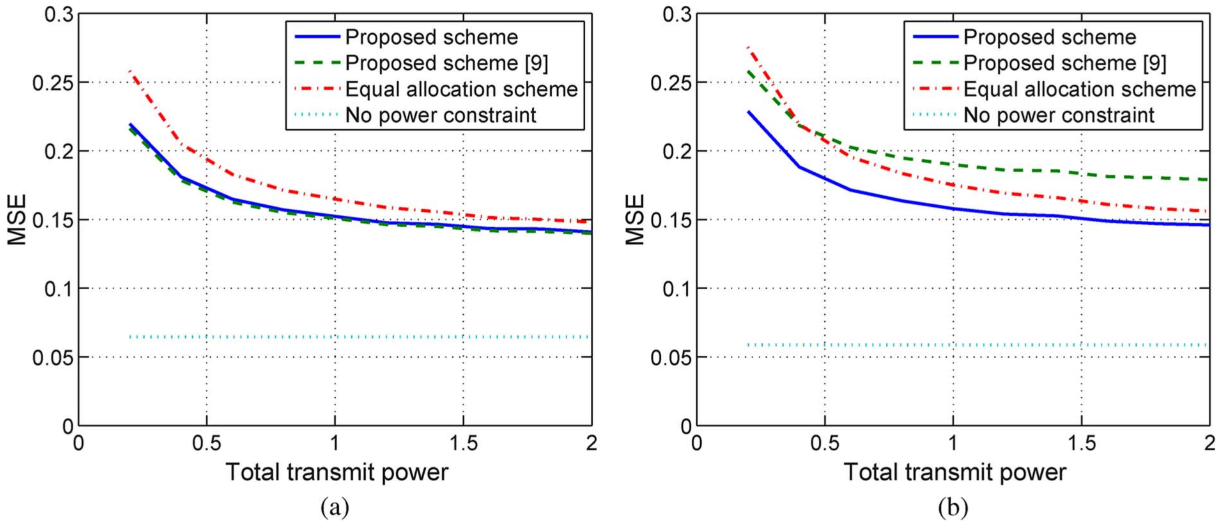

Fig. 4. Example B: MSEs versus total transmit power with uncorrelated and correlated observation noise, respectively. (a) $\beta = 0$; (b) $\beta = -0.9$.

distortion target, say, 0.08, the power required by our proposed method is about 2/3 of that required by [8] and 1/2 of that required by the equal power allocation scheme.

To gain insight into the power allocation in the presence of correlation, we consider the following illustrative example. We still assume linear model: $x_n = \theta + w_n$, where we set $N = 5$, and the observation noise $\{w_n\}$ is highly correlated with autocorrelation $\sigma_{w_i}^2 = 1\ \forall i$ and cross-correlation $\sigma_{w_i,w_j} = 0.95\ \forall i \neq j$. The channel SNRs are specified as follows: $\mathrm{SNR}_{\mathrm{ch},n} = 16$ dB for $n = 1, 5$ and $\mathrm{SNR}_{\mathrm{ch},n} = 3$ dB for $n = 2, 3, 4$. Fig. 2 is a bar graph depicting the power allocation given by the three schemes. We observe that our scheme assigns all the power to the sensors with the best-quality channels. This makes sense because, due to the high spatial correlation, sensors provide redundant information, and little data diversity can be exploited. In this case, a meaningful power allocation strategy is to allocate most power to the sensors with the best-quality channels. In contrast, by ignoring the spatial correlation among sensors, a waterfilling scheme [8] assigns power only according to the channel quality. As a result, some highly correlated sensors with poor channel quality but still above waterfilling threshold are assigned a considerable portion of the total power, as shown in Fig. 2. This, as indicated earlier, is unreasonable since highly correlated data provide little useful diversity.

We next consider a nonlinear model: $x_n = f_n(\theta) + w_n$, where we define $f_1(x) = f_2(x) = (1/3)x$, $f_3(x) = f_4(x) = \sin(x)$, and $f_5(x) = f_6(x) = \sin(2x)$; θ is a uniform random variable defined on the interval $(-\pi, \pi)$; the observation noise $\{w_n\}$ is independent with covariance matrix $\mathbf{R}_w = 0.01\mathbf{I}$. The wireless channels are generated according to (22). We compare our method with [8] and the equal power allocation scheme. The results are plotted in Fig. 3. Note that to apply [8] to the current scenario, [8] is implemented by ignoring the nonlinearity of the sensor observation model and treating it as if it were linear, i.e., the covariance matrix $\mathbf{R}_x$ in the objective function is replaced by $(1/\sigma_\theta^2)\mathbf{R}_{\theta x}^T\mathbf{R}_{\theta x} + \mathbf{R}_w$. Fig. 3 shows that, to attain a same estimation performance, considerable power savings can be achieved by our proposed method as compared to the other two schemes.

B. Estimation of a Random Field

Distributed estimation of a random field with a set of dependent parameters $\{\theta_n\}_{n=1}^N$ is studied. We consider a linear signal-plus-noise model: $x_n = \theta_n + w_n$. A distance-dependent function $E[\theta_i\theta_j] = \rho^{d_{i,j}}$ is employed to model the correlation among $\{\theta_n\}$, where $d_{i,j}$ denotes the Euclidean distance between sensors i and j. We simulate a moderately high correlation among the random parameters $\{\theta_n\}$, in which we choose $\rho = 0.9$ and $N = 5$, and the sensors are placed uniformly at random on a 3 × 3 area. The qualities of the channels are assumed unbalanced with $\mathrm{SNR}_{\mathrm{ch},n} = 15$ dB for $n = 1$, $\mathrm{SNR}_{\mathrm{ch},n} = 20$ dB for $n = 2, 5$ and $\mathrm{SNR}_{\mathrm{ch},n} = -20$ dB for $n = 3, 4$. For simplicity, the observation noise $\{w_n\}$ is, again, modeled as an AR(1) process (21) with $\sigma_u^2 = 0.1$, $\beta = 0$ and $\beta = -0.9$, respectively. Fig. 4 shows the distortion performance of the respective schemes, where the distortion performance is measured by the average MSE, i.e., $(1/N)E[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2]$. The results are averaged over 500 Monte Carlo runs, with the distribution of the sensors on the specified area independently generated for each run. The method [9] is optimal when $\beta = 0$, which corresponds to a scenario of independent observation noise. In this case, the method [9] performs the best among the three schemes [see Fig. 4(a)]. Nevertheless, we see that our proposed method performs quite similarly as the optimal one [9]. Also, as previous example, when the observation noise becomes correlated, i.e., $\beta = -0.9$, our method surpasses the other two schemes and a considerable power saving can be achieved.

VI. Conclusion

We studied a power allocation problem arising from distributed estimation of a random parameter in the presence of noisy links. A suboptimal solution was developed by maximizing a lower bound of the original optimum power allocation criterion. The proposed method can handle a general scenario where the sensor observations are nonlinear functions of the parameter of interest and the observation noise is spatially correlated. Numerical results show that our proposed method provides an efficient power assignment by taking into account sensor correlation and channel disparity. Considerable power savings can be achieved at the same distortion level as compared with other existing strategies.

References

[1] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.

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

[3] H. Li and J. Fang, "Distributed adaptive quantization and estimation for wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 669–672, Oct. 2007.

[4] K. Zhang, X. R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion—Part VI: Sensor data compression," presented at the Int. Conf. Information Fusion, Queensland, Australia, 2003.
[5] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1631–1639, May 2005.
[6] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4284–4299, Aug. 2007.
[7] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 413–422, Feb. 2006.
[8] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, no. 9, pp. 4683–4695, Sep. 2007.
[9] I. Bahceci and A. K. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Trans. Commun.*, vol. 56, no. 7, pp. 1146–1156, Jul. 2008.
[10] Y. Huang and Y. Hua, "Multi-hop progressive decentralized estimation in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 1004–1007, Dec. 2007.
[11] 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 Process.*, vol. 56, no. 5, pp. 2171–2176, May 2008.
[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.