Optimal Linear Fusion for Distributed Detection Via Semidefinite Programming

Zhi Quan, Wing-Kin Ma, Shuguang Cui, and Ali H. Sayed, *Fellow, IEEE*

***Abstract*—Consider the problem of signal detection via multiple distributed noisy sensors. We study a linear decision fusion rule of [Z. Quan, S. Cui, and A. H. Sayed, "Optimal Linear Cooperation for Spectrum Sensing in Cognitive Radio Networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–40, Feb. 2008] to combine the local statistics from individual sensors into a global statistic for binary hypothesis testing. The objective is to maximize the probability of detection subject to an upper limit on the probability of false alarm. We propose a more efficient solution that employs a divide-and-conquer strategy to divide the decision optimization problem into two subproblems. Each subproblem is a nonconvex program with a quadratic constraint. Through a judicious reformulation and by employing a special matrix decomposition technique, we show that the two nonconvex subproblems can be solved by semidefinite programs in a globally optimal fashion. Hence, we can obtain the optimal linear fusion rule for the distributed detection problem. Compared with the likelihood-ratio test approach, optimal linear fusion can achieve comparable performance with considerable design flexibility and reduced complexity.**

***Index Terms*—Distributed detection, hypothesis testing, nonconvex optimization, semidefinite programming.**

I. Introduction

Distributed detection techniques that use processed measurements from multiple spatially distributed sensors have a wide variety of applications in military surveillance, environmental monitoring, and wireless communications. In a distributed detection system, multiple sensors work collaboratively to distinguish between two or more hypotheses [3], [4]. Specifically, each sensor compresses its observations into a local statistic and then sends this information to a fusion center, which is responsible for making the final decision. A distinct feature that makes distributed detection challenging is that local observations need to be compressed individually before they are jointly processed by the fusion center. This feature is due to the large volume of data observed at local sensors as well as the limited channel capacity between the fusion center and sensors.

Manuscript received September 23, 2009; accepted November 29, 2009. First published December 31, 2009; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Pierre Vandergheynst. This research was supported in part by the National Science Foundation under Grants ECS-0601266, ECS-0725441, CCF-094936, CNS-0721935, and CCF-0726740, by the Department of Defense under Grant HDTRA-07-1-0037, and by a General Research Fund granted by the Hong Kong Research Grant Council (Project No. CUHK415908). This work was performed when Z. Quan was with the Electrical Engineering Department at UCLA. This work was presented in part at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Taipei, Taiwan, April 2009.

Z. Quan is with the Research and Development Division, Qualcomm Inc., San Diego, CA 92121 USA (e-mail: zquan@qualcomm.com).

W.-K. Ma is with the Department of Electronic Engineering, Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: wkma@ee.cuhk.edu.hk).

S. Cui is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: cui@ece.tamu.edu).

A. H. Sayed is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.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.2009.2039823

The likelihood-ratio test (LRT) is an optimal fusion rule for both hard and soft decision combining techniques according to the Neyman–Pearson criterion. A large portion of the literature on distributed detection has focused on the hard-decision fusion where each local sensor makes a local decision via the LRT and then sends the binary decision to the fusion center for global decision making [3]–[5]. Finding the local optimal decision rules is difficult since the thresholds at local sensors are coupled with each other. For independent observations, the optimality of the LRT at local sensors has been established [5], [6]; while computing the optimal local thresholds for correlated observations is an NP-complete problem [7], [8], one has to turn to suboptimal solutions [9]. To circumvent the need for optimizing local thresholds, each sensor can instead send a soft decision (summary statistic) to the fusion center in which only one optimal test needs to be performed.

In this correspondence, we study a linear fusion rule for distributed detection. The linear fusion rule is motivated by at least two advantages. First, linear fusion has less computational complexity, i.e., $O(n)$, than that of the quadratic LRT detector, i.e., $O(n^2)$. Second, the performance and the threshold of a linear fusion rule can be obtained numerically while the computation of the threshold and the performance of an LRT are mathematically intractable. To find the threshold for the expected performance of an LRT detector, one usually has to use Monte Carlo simulation, which is time consuming when the error probability is less than 10^{-5}. The linear fusion rule has been proposed and optimized by using the bisection search to solve a sequence of quadratically constrained quadratic programs (QCQPs) in the context of cognitive radio design in [2]. Here, we consider a more general case—the optimal design of linear fusion for distributed detection. We propose a fast algorithm to solve for the optimal solution via semidefinite programming (SDP). In essence, we show that the optimal linear fusion rule can be characterized by a rank-one solution of the formulated SDP problem.

Notation: $\mathbf{S}^n$ denotes the set of $n \times n$ symmetric matrices; $\mathrm{tr}(\cdot)$ denotes the trace of a matrix; $\det(\cdot)$ denotes the determinant of a matrix; $\mathrm{diag}(\cdot)$ denotes a diagonal matrix; $Q(\cdot)$ denotes the *complementary cumulative distribution* function, i.e., $Q(x) = 1/\sqrt{2\pi}\int_x^{+\infty} e^{-\tau^2/2} d\tau$; $\succ$ ($\succeq$) denotes the matrix inequality, i.e., $\mathbf{A} \succ (\succeq) \mathbf{B}$ signifies that $\mathbf{A} - \mathbf{B}$ is positive definite (semidefinite).

II. System Model

Consider a network of N distributed sensors, each of which is observing a phenomenon under the two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. Let Y_i denote the observation at the ith sensor. Each sensor employs the mapping rule $u_i = g_i(Y_i)$ and transmits u_i to the fusion center. Based on the received information $\mathbf{u} = [u_1, u_2, \ldots, u_N]^T$, the fusion center makes the global decision on one of the two hypotheses.

Suppose that the received vector $\mathbf{u}$ at the fusion center can be treated as a realization generated from an N-dimensional normal (Gaussian) distribution under each hypothesis, i.e.,

$$\mathbf{u} \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), & \mathcal{H}_0 \\ \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), & \mathcal{H}_1 \end{cases} \tag{1}$$

where $\boldsymbol{\mu}_0(\boldsymbol{\mu}_1)$ and $\boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_1)$ are the mean vector and covariance matrix of $\mathbf{u}$ under $\mathcal{H}_0(\mathcal{H}_1)$. Note that $\boldsymbol{\Sigma}_0 \succeq \mathbf{0}$ and $\boldsymbol{\Sigma}_1 \succeq \mathbf{0}$. This model is useful, for example, in radio astronomy, sensor networks, and other applications in which the background noise is normally distributed according to the central limit theorem; see [2] on how such a model can be motivated in a practical distributed detection system. Hence, the probability distribution function (pdf) of $\mathbf{u}$ can be written as

$$p(\mathbf{u};\mathcal{H}_i) = \frac{1}{(2\pi)^{N/2}\det^{1/2}(\boldsymbol{\Sigma}_i)} \times \exp\left[-\frac{1}{2}(\mathbf{u}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{u}-\boldsymbol{\mu}_i)\right] \quad (2)$$

where $i = 0, 1$. The detection performance can be evaluated in terms of the probability of false alarm, $P_f = \Pr(\mathcal{H}_1 \text{ is decided}|\mathcal{H}_0)$, and the probability of detection, $P_d = \Pr(\mathcal{H}_1 \text{ is decided}|\mathcal{H}_1)$. According to the Neyman–Pearson criterion, an optimal detector will maximize P_d with an upper limit on P_f, i.e.,

$$\max \quad P_d \quad \text{s.t.} \quad P_f \le \varepsilon. \quad (3)$$

A. Likelihood-Ratio Test

LRT has been shown to be the optimal fusion rule [10], which is obtained by first computing the likelihood ratio

$$\frac{p(\mathbf{u};\mathcal{H}_1)}{p(\mathbf{u};\mathcal{H}_0)} = \frac{\det^{1/2}(\boldsymbol{\Sigma}_0)}{\det^{1/2}(\boldsymbol{\Sigma}_1)}\exp \times \left[\frac{1}{2}\mathbf{u}^T\left(\boldsymbol{\Sigma}_0^{-1}-\boldsymbol{\Sigma}_1^{-1}\right)\mathbf{u} + \left(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\right)\mathbf{u}\right] \quad (4)$$

and then taking the logarithm, and comparing with a decision threshold:

$$\Lambda(\mathbf{u}) = \mathbf{u}^T\left(\boldsymbol{\Sigma}_0^{-1}-\boldsymbol{\Sigma}_1^{-1}\right)\mathbf{u} + 2\left(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}-\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\right)\mathbf{u} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_{\mathrm{LRT}}. \quad (5)$$

Since $\Lambda(\mathbf{u})$ has a quadratic form, the evaluation of its pdf requires multidimensional integrations, and potentially involves complicated decision regions. Consequently, the analysis of the detection performance (P_f, P_d) and the choice of the optimal threshold γ_{LRT} are generally mathematically intractable.

B. Linear Fusion

To circumvent this difficulty, we propose a linear fusion rule

$$T(\mathbf{u}) = \sum_{i=1}^{N} w_i u_i = \mathbf{w}^T\mathbf{u} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \quad (6)$$

where $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T$ are the weight coefficients. Since the linear combination of multiple Gaussian random variables is still Gaussian, it can be verified that

$$T(\mathbf{u}) \sim \begin{cases} \mathcal{N}\left(\mathbf{w}^T\boldsymbol{\mu}_0, \mathbf{w}^T\boldsymbol{\Sigma}_0\mathbf{w}\right), & \mathcal{H}_0 \\ \mathcal{N}\left(\mathbf{w}^T\boldsymbol{\mu}_1, \mathbf{w}^T\boldsymbol{\Sigma}_1\mathbf{w}\right), & \mathcal{H}_1. \end{cases} \quad (7)$$

Accordingly, the probabilities of false alarm and detection can be expressed as

$$P_f = P(T(\mathbf{u}) \ge \gamma|\mathcal{H}_0) = Q\left(\frac{\gamma - \mathbf{w}^T\boldsymbol{\mu}_0}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_0\mathbf{w}}}\right) \quad (8)$$

and

$$P_d = P(T(\mathbf{u}) \ge \gamma|\mathcal{H}_1) = Q\left(\frac{\gamma - \mathbf{w}^T\boldsymbol{\mu}_1}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_1\mathbf{w}}}\right). \quad (9)$$

Our objective is to find the optimal weight vector $\mathbf{w}$ that maximizes P_d subject to some constraint on P_f.

III. Semi-Definite Programming Formulation

In this section, we show how to optimize the linear fusion rule (6) via SDP formulation. From (8), we first express γ as a function of the required probability of false alarm (by setting $P_f = \varepsilon$) and the weight coefficients $\mathbf{w}$:

$$\gamma = \mathbf{w}^T\boldsymbol{\mu}_0 + Q^{-1}(\varepsilon)\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_0\mathbf{w}}. \quad (10)$$

Plugging (10) into (9) gives an unconstrained optimization problem:

$$\max_{\mathbf{w}} P_d = Q\left[\frac{Q^{-1}(\varepsilon)\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_0\mathbf{w}} - \mathbf{w}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_1\mathbf{w}}}\right]. \quad (11)$$

Since $Q(\cdot)$ is a monotonically non-increasing function, maximizing P_d is equivalent to minimizing the term inside the Q-function in (11). Consequently, (3) can be converted into an equivalent form:

$$\min_{\mathbf{w}} f(\mathbf{w}) = \frac{Q^{-1}(\varepsilon)\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_0\mathbf{w}} - (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{w}}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_1\mathbf{w}}}. \quad (12)$$

Solving (12) is difficult since it is a nonconvex problem. To find the globally optimal solution, we can employ a divide-and-conquer strategy to exploit the special problem structure. Hence, we solve this nonconvex optimization problem via tackling the following two subproblems.[1]

A. $P_d \le 1/2$

First, consider the case in which $P_d \le 1/2$, i.e., $f(\mathbf{w}) \ge 0$. It has been shown in [2] that in this case problem (12) is equivalent to

$$\begin{aligned} \min_{\mathbf{z}} \quad & Q^{-1}(\varepsilon)\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z}} - (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{z} \\ \text{s.t.} \quad & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \ge 1 \end{aligned} \quad (13)$$

where $\mathbf{z} = \mathbf{w}/\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}_1\mathbf{w}}$. By introducing a new variable $\alpha = Q^{-1}(\varepsilon)\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z}} - (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{z}$, we can transform (13) into

$$\begin{aligned} \min_{\mathbf{z}} \quad & \alpha^2 \\ \text{s.t.} \quad & \alpha = Q^{-1}(\varepsilon)\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z}} - (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{z},\ \alpha \ge 0 \\ & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \ge 1 \end{aligned} \quad (14)$$

where we utilize the fact that minimizing α^2 is equivalent to minimizing α when α is nonnegative. If the above problem is feasible, we will be able to solve (14) and hence (12); otherwise, we turn to use the problem formulation in (22) for Case B. We show in Appendix A that (14) is equivalent to

$$\begin{aligned} \min_{\mathbf{z}} \quad & \alpha^2 \\ \text{s.t.} \quad & \left[\alpha + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\mathbf{z}\right]^2 = Q^{-2}(\varepsilon)\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z}, \quad \alpha \ge 0 \\ & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \ge 1 \end{aligned} \quad (15)$$

which is a nonconvex quadratic program subject to two quadratic constraints and a linear constraint.

Let $\mathbf{x} = \left[\mathbf{z}^T\ \alpha\right]^T$. Problem (15) can be written as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^T\mathbf{F}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^T\mathbf{G}_1\mathbf{x} = 0, \quad \mathbf{x}^T\mathbf{H}\mathbf{x} \ge 1 \end{aligned} \quad (16)$$

[1]Please note that to apply the proposed approach, we do not need to know whether the resulting detection probability is below or above 1/2. A simple method is to solve both subproblems, within which the better solution will be the exactly optimal.

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{0}^{N\times N} & \mathbf{0}^{N\times 1} \\ \mathbf{0}^{1\times N} & 1 \end{bmatrix} \tag{17}$$

$$\mathbf{G}_1 = \begin{bmatrix} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T - Q^{-2}(\varepsilon)\boldsymbol{\Sigma}_0 & \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 \\ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T & 1 \end{bmatrix} \tag{18}$$

and

$$\mathbf{H} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0}^{N\times 1} \\ \mathbf{0}^{1\times N} & 0 \end{bmatrix} \tag{19}$$

with $\mathbf{F}$, $\mathbf{G}_1$, and $\mathbf{H} \in \mathbf{S}^{N+1}$. Since $\mathbf{F} \succeq \mathbf{0}$ and $\mathbf{H} \succeq \mathbf{0}$, an optimal solution of (16) must satisfy $\mathbf{x}^T\mathbf{H}\mathbf{x} = 1$. Thus, (16) is equivalent to

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^T\mathbf{F}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^T\mathbf{G}_1\mathbf{x} = 0, \quad \mathbf{x}^T\mathbf{H}\mathbf{x} = 1. \end{aligned} \tag{20}$$

Applying SDP relaxation [11] with the hidden rank-one constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^T \in \mathbf{S}^{N+1}$ eliminated, we obtain a standard SDP problem as follows:

$$\begin{aligned} \min_{\mathbf{X}\in\mathbf{S}^{N+1}} \quad & \mathrm{tr}\,(\mathbf{F}\mathbf{X}) \\ \text{s.t.} \quad & \mathrm{tr}\,(\mathbf{G}_1\mathbf{X}) = 0, \quad \mathrm{tr}\,(\mathbf{H}\mathbf{X}) = 1 \\ & \mathbf{X} \succeq \mathbf{0} \end{aligned} \tag{21}$$

which has linear equality constraints and a matrix nonnegativity constraint on the unknown matrix $\mathbf{X}$. Recall that $\mathrm{tr}\,(\mathbf{F}\mathbf{X}) = \sum_{i=1}^{N+1}\sum_{j=1}^{N+1} F_{ij}X_{ji}$ is in the form of a generalized real-valued linear function on $\mathbf{S}^{N+1}$, which shows that SDP is a generalized framework of linear programming over matrices.

As a result, (21) is a relaxation of (20) since we have removed the rank-one constraint. We will show in Section IV that there exists at least one rank-one solution for (21) such that its optimal value is equal to the globally optimal value of (20). Hence, this rank-one solution is also the optimal solution of (16).

B. $P_d > 1/2$

For the case where $P_d > 1/2$, i.e., $f(\mathbf{w}) < 0$, problem (12) can be written as

$$\begin{aligned} \max_{\mathbf{z}} \quad & -Q^{-1}(\varepsilon)\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z}} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\mathbf{z} \\ \text{s.t.} \quad & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \le 1 \end{aligned} \tag{22}$$

by which we could also obtain a positive objective function. By defining $\alpha = -Q^{-1}(\varepsilon)\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z}} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\mathbf{z}$ in a similar way as in Case A, we can transform (22) into an equivalent form as

$$\begin{aligned} \max_{\mathbf{z}} \quad & \alpha^2 \\ \text{s.t.} \quad & \left[\alpha - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\mathbf{z}\right]^2 = Q^{-2}(\varepsilon)\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z},\ \alpha > 0 \\ & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \le 1. \end{aligned}$$

Defining the matrix

$$\mathbf{G}_2 = \begin{bmatrix} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T - Q^{-2}(\varepsilon)\boldsymbol{\Sigma}_0 & -(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \\ -(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T & 1 \end{bmatrix} \tag{23}$$

and considering (17)–(19), we can show that (23) is equivalent to

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^T\mathbf{F}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^T\mathbf{G}_2\mathbf{x} = 0, \quad \mathbf{x}^T\mathbf{H}\mathbf{x} = 1. \end{aligned} \tag{24}$$

Like (21), the SDP relaxation of (24) is given by

$$\begin{aligned} \max_{\mathbf{X}\in\mathbf{S}^{N+1}} \quad & \mathrm{tr}\,(\mathbf{F}\mathbf{X}) \\ \text{s.t.} \quad & \mathrm{tr}\,(\mathbf{G}_2\mathbf{X}) = 0, \quad \mathrm{tr}\,(\mathbf{H}\mathbf{X}) = 1 \\ & \mathbf{X} \succeq \mathbf{0} \end{aligned} \tag{25}$$

whose rank-one solution solves (22). In particular, the way to construct the desired rank-one solution will be discussed in the following section. Throughout the rest of the correspondence, we assume that the problems satisfy the Slater's regularity condition [11].

IV. Rank-One Solution

The rank-one solutions for our particular problems in (21) and (25) can be obtained by employing a special rank-one decomposition technique proposed in [12].

Lemma 1: Let $\mathbf{X} \in \mathbf{S}^n, \mathbf{X} \succeq \mathbf{0}$ be a matrix with rank r. Given an arbitrary matrix $\mathbf{G} \in \mathbf{S}^n$, $\mathbf{X}$ can be decomposed into

$$\mathbf{X} = \sum_{i=1}^{r} \mathbf{x}_i\mathbf{x}_i^T \tag{26}$$

where the decomposed vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ satisfy

$$\mathbf{x}_i^T\mathbf{G}\mathbf{x}_i = \frac{\mathrm{tr}(\mathbf{G}\mathbf{X})}{r}, \quad i = 1, 2, \ldots, r. \tag{27}$$

■

Interested readers are referred to [12] for the proof of Lemma 1. From the proof one can see the rank-one decomposition procedure, which we summarize in Algorithm 1. This matrix decomposition technique has been used in [12] to solve nonconvex quadratic programs with two quadratic constraints. However, the method in [12] is not directly applicable to solving the problems (21) and (25). In the sequel, we will show how to develop the optimal rank-one solutions for (21) and (25).

Algorithm 1: Matrix Decomposition Procedure

Input: $\mathbf{X} \succeq \mathbf{0}$, and $\mathbf{G} \in \mathbf{S}^n$.

Step 1: Apply any decomposition that yields

$$\mathbf{X} = \sum_{i=1}^{r} \mathbf{x}_i\mathbf{x}_i^T, \quad \text{e.g., eigendecomposition.}$$

Step 2: If $\mathbf{x}_i^T\mathbf{G}\mathbf{x}_i = \mathrm{tr}(\mathbf{G}\mathbf{X})/r$ for all i then output $\mathbf{x}_1, \ldots, \mathbf{x}_r$ and return; otherwise find i, j such that $\mathbf{x}_i^T\mathbf{G}\mathbf{x}_i > \mathrm{tr}(\mathbf{G}\mathbf{X})/r$ and $\mathbf{x}_j^T\mathbf{G}\mathbf{x}_j < \mathrm{tr}(\mathbf{G}\mathbf{X})/r$.

Step 3: Determine β such that

$$(\mathbf{x}_i + \beta\mathbf{x}_j)^T\mathbf{G}(\mathbf{x}_i + \beta\mathbf{x}_j) = \frac{(1+\beta^2)\,\mathrm{tr}(\mathbf{G}\mathbf{X})}{r}.$$

Step 4: $\mathbf{x}_i := (\mathbf{x}_i + \beta\mathbf{x}_j)/\sqrt{1+\beta^2}$,

$$\mathbf{x}_j := \frac{(-\beta\mathbf{x}_i + \mathbf{x}_j)}{\sqrt{1+\beta^2}}.$$

Step 5: Repeat Step 2.

In this algorithm, it can be shown by contradiction that there always exist an underfit vector and an overfit vector in pairs. It is also indicated in [12] that convergence to the desired condition (27) is guaranteed for a finite number of iterations.

Theorem 1: There exists at least one rank-one matrix that optimizes (21).

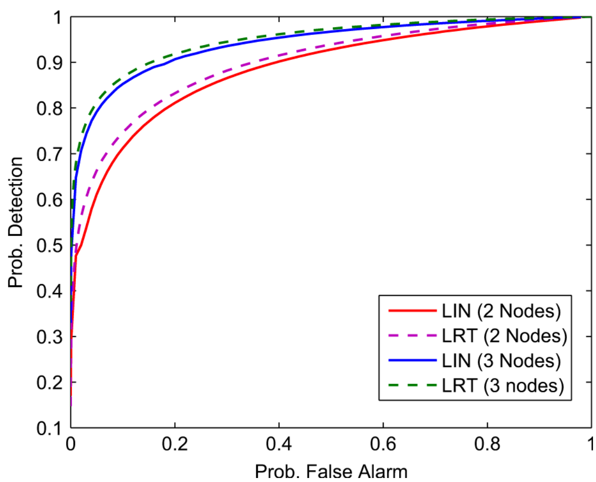

Fig. 1. The ROC curves of linear fusion and LRT. For two nodes, $\boldsymbol{\mu}_0 = [0, 0]$, $\boldsymbol{\mu}_1 = [1, 2]$, $\boldsymbol{\Sigma}_0 = [1, 0; 0, 1]$, and $\boldsymbol{\Sigma}_1 = [1, 1/2; 1/2, 1]$. For three nodes, $\boldsymbol{\mu}_0 = [0, 0, 0]$, $\boldsymbol{\mu}_1 = [1, 2, 2]$, $\boldsymbol{\Sigma}_0 = [1, 0, 0; 0, 1, 0; 0, 0, 1]$, and $\boldsymbol{\Sigma}_1 = [1, 1/2, 0; 1/2, 1, 1/2; 0, 1/2, 1]$.

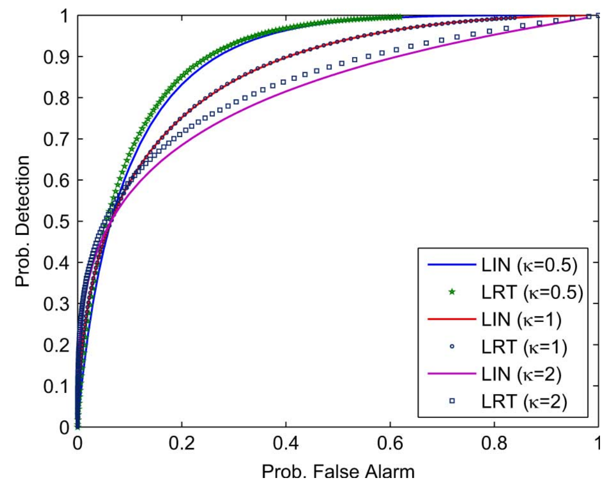

Fig. 2. The ROC curves with $\Sigma_0 = \mathrm{diag}([1, 1])$ and $\Sigma_1 = \kappa\Sigma_0$, where κ implies the difference between $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$.

Proof: The proof is given in Appendix B. ■

Specifically, we give the following theorem to construct the optimal rank-one solutions for (21).

Theorem 2: Let $\bar{\mathbf{X}}$ be an optimal solution to the SDP relaxation problem (21) with rank r. A rank-one solution to (21) can be obtained from $\bar{\mathbf{X}}$ by performing the following steps.

i) Apply the matrix decomposition on $\bar{\mathbf{X}}$ with respect to $\mathbf{G}_1$ to obtain vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$.

ii) For any l satisfying $\mathbf{x}_l^T \mathbf{H} \mathbf{x}_l > 0, 1 \le l \le r$, let

$$\bar{\mathbf{x}}_l = \frac{\mathbf{x}_l}{\sqrt{\mathbf{x}_l^T \mathbf{H} \mathbf{x}_l}}.$$

We have $\bar{\mathbf{x}}_l \bar{\mathbf{x}}_l^T$ as the optimal rank-one solution for (21).

Proof: Theorem 2 is proven by showing that the rank-one matrix $\bar{\mathbf{x}}_l \bar{\mathbf{x}}_l^T$ satisfies the Karush–Kuhn–Tucker (KKT) optimality conditions [11] of (21). The detailed proof is given in Appendix C. ■

We shall note that the principles in Theorems 1 and 2 are perfectly applicable to solving the SDP problem in (25) by replacing $\mathbf{G}_1$ with $\mathbf{G}_2$.

V. Numerical Results

We now evaluate the performance of the optimal linear detector. In Fig. 1, we illustrate the receiver operating characteristic (ROC) of the optimal linear detector (denoted by LIN) over various SNR levels, where the LRT detector serves as a performance upper bound. We see that the optimal linear detector approaches the LRT performance limit as the number of sensors increases because the effective SNR at the fusion center increases. This observation is due to the exponential fall-off of the Q function. In Fig. 2, we show how the difference between $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ affects the detection performance. It can be observed that the optimal linear detector approximates the LRT detector well if the difference between $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ is small (e.g., $\kappa \approx 1$). In the special case where $\kappa = 1$, the LRT detector degenerates into a linear detector, as seen in (5). On the other hand, the linear detector might perform far away from the optimum if the difference between the two covariance matrices is large.

VI. Conclusion

We have studied a linear fusion strategy for distributed detection, which can be optimized through SDP reformulation. Compared with the optimal LRT detector, the linear detector achieves comparable performance with reduced complexity under conditions of reasonable SNR levels or small difference between the covariance matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$. These conditions provides useful insights for practicing engineers to trade performance with complexity in designing distributed detection systems.

Appendix A
Equivalence of (14) and (15)

We find that (15) can be rewritten as $\bar{f} = \min\{\bar{f}_1, \bar{f}_2\}$, where

$$\begin{aligned} \bar{f}_1 = \min_{\mathbf{z}} \quad & \alpha^2 \\ \text{s.t.} \quad & \alpha + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{z} = Q^{-1}(\varepsilon)\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z} \\ & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \ge 1 \end{aligned} \tag{28}$$

and

$$\begin{aligned} \bar{f}_2 = \min_{\mathbf{z}} \quad & \alpha^2 \\ \text{s.t.} \quad & \alpha + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{z} = -Q^{-1}(\varepsilon)\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z} \\ & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \ge 1, \end{aligned} \tag{29}$$

both under the condition $\alpha \ge 0$. Note that (28) is exactly the same as (14). Furthermore, we have

$$\begin{aligned} \bar{f}_2 = \min_{\mathbf{z}} \quad & \left[Q^{-1}(\varepsilon)\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}_0\mathbf{z}} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{z}\right]^2 \\ \text{s.t.} \quad & \mathbf{z}^T\boldsymbol{\Sigma}_1\mathbf{z} \ge 1 \end{aligned} \tag{30a}$$

$$\begin{aligned} = \min_{\mathbf{y}} \quad & \left[Q^{-1}(\varepsilon)\sqrt{\mathbf{y}^T\boldsymbol{\Sigma}_0\mathbf{y}} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \mathbf{y}\right]^2 \\ \text{s.t.} \quad & \mathbf{y}^T\boldsymbol{\Sigma}_1\mathbf{y} \ge 1 \\ = \bar{f}_1 & \end{aligned} \tag{30b}$$

where (30b) is obtained via a change of variables, i.e., $\mathbf{z} = -\mathbf{y}$. Therefore, we conclude that (14) and (15) are equivalent.

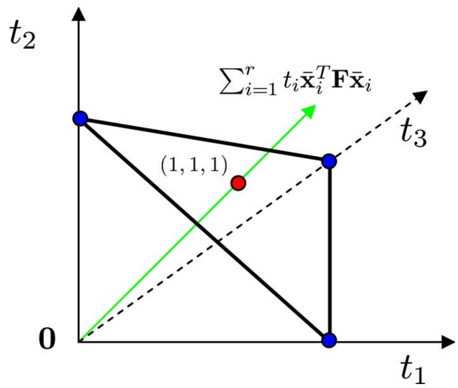

Fig. 3. The basic optimal feasible solution of the linear program (33) takes one of the three vertices, where the constraints constructs a polyhedron. In particular, the optimal extreme point is confined to the hyperplane $\sum_{i=1}^{r} t_i \bar{\mathbf{x}}_i^T \mathbf{H} \bar{\mathbf{x}}_i = 1$.

APPENDIX B

Proof: Suppose that $\bar{\mathbf{X}}$ is a minimizer of (21) with rank $r = \mathrm{rank}(\bar{\mathbf{X}})$. From Lemma 1, $\bar{\mathbf{X}}$ can be decomposed with respect to $\mathbf{G}_1$ as

$$\bar{\mathbf{X}} = \sum_{i=1}^{r} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T. \quad (31)$$

Since $\mathrm{tr}\left(\mathbf{G}_1 \bar{\mathbf{X}}\right) = 0$, we can infer from Lemma 1 that

$$\mathrm{tr}\left(\mathbf{G}_1 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T\right) = 0, \quad i = 1, 2, \ldots, r. \quad (32)$$

Consider a linear program

$$\begin{aligned} \min_{\mathbf{t}} \quad & \sum_{i=1}^{r} t_i \bar{\mathbf{x}}_i^T \mathbf{F} \bar{\mathbf{x}}_i \\ \text{s.t.} \quad & \sum_{i=1}^{r} t_i \bar{\mathbf{x}}_i^T \mathbf{H} \bar{\mathbf{x}}_i = 1 \\ & t_i \geq 0, \quad i = 1, 2, \ldots, r \end{aligned} \quad (33)$$

where $\mathbf{t} = [t_1, t_2, \ldots, t_r]^T$. Note that $t_1 = t_2 = \cdots = t_r = 1$ is an optimal solution for (33) since for any $t_i \geq 0, i = 1, 2, \ldots, r$,

$$\mathbf{X} = \sum_{i=1}^{r} t_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \quad (34)$$

is a feasible solution for (21). Thus, the minimal objective value of (21) is equal to that of (33), which corresponds to $t_1 = t_2 = \cdots = t_r = 1$.

Since the linear program (33) is bounded, it must have a basic optimal feasible solution (or an optimal extreme point), at which we should have at least r active constraints (including the equality constraint). The geometric illustration is given in Fig. 3. Hence, we should have at most one inactive constraint since the total number of constraints of (33) is $r + 1$. Namely, one of the r variables $\{t_i\}_{i=1}^{r}$ is positive at the basic optimal solution. Suppose that the positive variable is t_p $(1 \leq p \leq r)$. Thus, the rank-one matrix

$$\mathbf{X}^* = t_p \bar{\mathbf{x}}_p \bar{\mathbf{x}}_p^T$$

is also an optimal solution for (21). ■

APPENDIX C

Proof: Consider the Lagrangian of (21):

$$L(\mathbf{X}, \mathbf{Z}, \boldsymbol{\nu}) = \mathrm{tr}\left[(\mathbf{F} + \nu_1 \mathbf{G}_1 + \nu_2 \mathbf{H} - \mathbf{Z})\, \mathbf{X}\right] - \nu_2 \quad (35)$$

where $\mathbf{Z} \succeq \mathbf{0}$ and $\boldsymbol{\nu} = [\nu_1, \nu_2]^T$ are the dual variables or multipliers. The dual function is given by

$$\begin{aligned} g(\mathbf{Z}, \boldsymbol{\nu}) &= \inf_{\mathbf{X}} L(\mathbf{X}, \mathbf{Z}, \boldsymbol{\nu}) \\ &= \begin{cases} -\nu_2 & \mathbf{Z} = \mathbf{F} + \nu_1 \mathbf{G}_1 + \nu_2 \mathbf{H} \\ -\infty & \text{otherwise.} \end{cases} \end{aligned} \quad (36)$$

Consequently, the dual problem of (21) can be shown to be

$$\max_{\mathbf{Z}, \boldsymbol{\nu}} -\nu_2 \quad \text{s.t.} \quad \mathbf{Z} = \mathbf{F} + \nu_1 \mathbf{G}_1 + \nu_2 \mathbf{H}\ \mathbf{Z} \succeq \mathbf{0} \quad (37)$$

where $\mathbf{Z}$ and $\boldsymbol{\nu}$ and are the corresponding dual variables. When (21) satisfies the Slater's condition [11], strong duality holds for the primary-dual pair (21) and (37). Hence, $(\bar{\mathbf{X}}, \bar{\mathbf{Z}}, \bar{\boldsymbol{\nu}})$ are primary-dual optimal if, and only if, the KKT conditions [11] are satisfied:

$$\bar{\mathbf{X}} \succeq \mathbf{0}, \ \mathrm{tr}\left(\mathbf{G}_1 \bar{\mathbf{X}}\right) = 0, \ \mathrm{tr}\left(\mathbf{H} \bar{\mathbf{X}}\right) = 1 \quad (38a)$$

$$\bar{\mathbf{Z}} \succeq \mathbf{0} \quad (38b)$$

$$\bar{\mathbf{Z}} = \mathbf{F} + \bar{\nu}_1 \mathbf{G}_1 + \bar{\nu}_2 \mathbf{H} \quad (38c)$$

$$\bar{\mathbf{Z}} \bar{\mathbf{X}} = \mathbf{0}. \quad (38d)$$

Suppose that we have obtained such a primary-dual optimal point $(\bar{\mathbf{X}}, \bar{\mathbf{Z}}, \bar{\boldsymbol{\nu}})$ (say, numerically by using an interior-point method). Apply the rank-one decomposition of $\bar{\mathbf{X}}$ with respect to $\mathbf{G}_1$:

$$\bar{\mathbf{X}} = \sum_{i=1}^{r} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \quad (39)$$

where $\bar{\mathbf{x}}_i^T \mathbf{G}_1 \bar{\mathbf{x}}_i = \mathrm{tr}(\mathbf{G}_1 \bar{\mathbf{X}}) = 0$ for $i = 1, 2, \ldots, r$. Then, we have $\mathrm{tr}\left(\mathbf{G}_1 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T\right) = 0$ for all i's. From (38b)–(38d), we can infer that

$$(\mathbf{F} + \bar{\nu}_1 \mathbf{G}_1 + \bar{\nu}_2 \mathbf{H})\, \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T = \mathbf{0} \quad (40)$$

for $1 \leq i \leq r$. Moreover, using the fact that $\mathbf{H} \succeq \mathbf{0}$, we have

$$1 = \mathrm{tr}\left(\mathbf{H} \bar{\mathbf{X}}\right) \geq \mathrm{tr}\left(\mathbf{H} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T\right) = \bar{\mathbf{x}}_i^T \mathbf{H} \bar{\mathbf{x}}_i.r. \quad (41)$$

From (38a), we know that there must exist an l such that $\bar{\mathbf{x}}_l^T \mathbf{H} \bar{\mathbf{x}}_l > 0$. Let

$$\hat{\mathbf{x}}_l = \frac{\bar{\mathbf{x}}_l}{\sqrt{\bar{\mathbf{x}}_l^T \mathbf{H} \bar{\mathbf{x}}_l}}. \quad (42)$$

We see that $(\hat{\mathbf{x}}_l \hat{\mathbf{x}}_l^T, \bar{\mathbf{Z}}, \bar{\boldsymbol{\nu}})$ satisfies the KKT conditions in (38a). ■

REFERENCES

[1] Z. Quan, W.-K. Ma, S. Cui, and A. H. Sayed, "Optimal linear fusion for distributed spectrum sensing via semidefinite programming," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3629–3632.

[2] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–40, Feb. 2008.

[3] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, Jan. 1997.

[4] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, pp. 64–79, Jan. 1997.

[5] S. Thomopoulos, R. Viswanathan, and D. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 761–765, Sep. 1989.

[6] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, pp. 24–30, Jan. 1987.

[7] R. S. Blum, "Necessary conditions for optimum distributed sensor detectors under the Neyman–Pearson criterion," *IEEE Trans. Inf. Theory*, vol. 42, pp. 990–994, May 1996.

[8] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Autom. Control*, vol. 30, pp. 440–446, May 1985.

[9] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 693–699, Feb. 2005.
[10] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 916–920, Jul. 1992.
[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[12] J. F. Sturm and S. Zhang, "On cones of nonnegative quadratic functions," *Math. Oper. Res.*, vol. 28, no. 2, pp. 246–267, May 2003.

On the Optimal Performance in Asymmetric Gaussian Wireless Sensor Networks With Fading

Hamid Behroozi, Fady Alajaji, and Tamás Linder

***Abstract*—We study the estimation of a Gaussian source by a Gaussian wireless sensor network (WSN) where L distributed sensors transmit noisy observations of the source through a fading Gaussian multiple access channel to a fusion center. In a recent work Gastpar, ["Uncoded transmission is exactly optimal for a Simple Gaussian Sensor Network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008] showed that for a symmetric Gaussian WSN with no fading, uncoded (analog) transmission achieves the optimal performance theoretically attainable (OPTA). In this correspondence, we consider an asymmetric fading WSN in which the sensors have differing noise and transmission powers. We first present lower and upper bounds on the system's OPTA under random fading. We next focus on asymmetric networks with deterministic fading. By comparing the obtained lower and upper OPTA bounds under deterministic fading, we provide a sufficient condition for the optimality of the uncoded transmission scheme for a given power tuple $\boldsymbol{P} = (P_1, P_2, \ldots, P_L)$. Then, allowing the sensor powers to vary under a weighted sum constraint (this includes the sum-power constraint as a special case), we obtain a sufficient condition for the optimality of uncoded transmission and provide the system's corresponding OPTA.**

***Index Terms*—Gaussian multiple access channel with fading, joint source-channel coding, power-distortion tradeoff, remote source coding, sensor networks, uncoded transmission.**

I. Introduction

We consider the estimation of a memoryless Gaussian source by a Gaussian wireless sensor network (WSN) where L sensors observe the source signal X corrupted by additive independent noise. The overall system is depicted in Fig. 1. The sensors communicate information about their observations through a fading Gaussian multiple access channel (MAC) to a single fusion center (FC). The fading coefficients are not known by the encoders but are available at the FC. The encoders are distributed and cannot cooperate to exploit the correlation between their inputs. Each encoder is subject to a transmission cost constraint. The FC aims to reconstruct the main source X at the smallest cost in the communication link. Our interest lies in determining the optimal power-distortion region, with the fidelity of estimation at the FC measured by the mean squared-error (MSE) distortion. Specifically, for a given L-tuple of sensor powers $\boldsymbol{P} = (P_1, P_2, \ldots, P_L)$, we seek to determine the system's minimum achievable distortion which we refer to as the optimal performance theoretically attainable (OPTA).

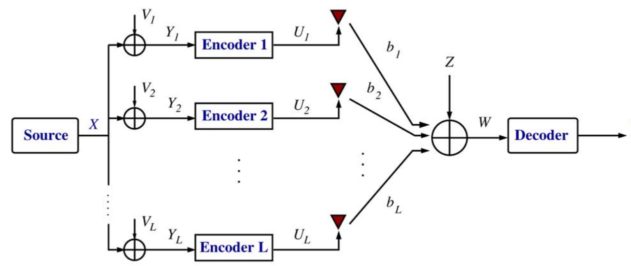

Fig. 1. A Gaussian wireless sensor network (WSN) with fading.

In [1] and [2], it is proved that uncoded transmission is exactly optimal for symmetric Gaussian WSNs with a finite number of sensors and no fading. Uncoded transmission in this case (and in the rest of this correspondence) means scaling the encoder input subject to the channel power constraint and transmitting without explicit channel coding. Note that the separate source and channel coding theorem of Shannon [3] does not hold for this problem [1], [2]. In the case of deterministic fading, lower and upper bounds on the minimum distortion are presented in [1], [2], and [4], and for random fading, bounds are also presented in [4] and [5]. The minimum achievable distortion under a sum-power constraint for the uncoded transmission scheme in the WSN with deterministic fading is presented in [6]. The optimality of uncoded transmission in some other multiuser communication systems was recently shown in [7] and [8].

For the asymmetric fading Gaussian WSN, the following important issues remain unknown: Under either random or deterministic fading, what is the system's OPTA? Also, What is the optimal coding strategy that achieves OPTA? Our main contributions in this correspondence are as follows: First, by applying the idea of maximum correlation coefficient, illustrated in [9]–[11], we generalize the OPTA lower bound in [1] to an asymmetric Gaussian WSN with random fading. We show that the new bound is a tighter lower bound on the OPTA than that of [5] for a Gaussian WSN with random fading. We also analyze the uncoded transmission scheme and provide an upper bound on the OPTA for a given set of sensor powers. These two bounds constitute an extension of the bounds given for deterministic fading case in [1] and [2]. We next specialize the results to the case of deterministic fading. We establish a condition under which the lower and upper bounds on the system's OPTA coincide, hence making the uncoded transmission scheme optimal. We next allow the sensor powers to vary under a linear combination of powers (LCP) constraint. Aside from being a natural generalization of the sum-power constraint, the LCP constraint explicitly allows to introduce weight coefficients that reflect the potentially differing costs of supplying power to individual sensors. Our final contribution is to provide sufficient conditions for the optimality of un-

Manuscript received June 28, 2009; accepted December 19, 2009. First published January 15, 2010; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Huaiyu Dai. This work was supported in part by a Postdoctoral Fellowship from the Ontario Ministry of Research and Innovation (MRI) and by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory (ISIT), Toronto, ON, Canada, July 2008, and at the IEEE Canadian Workshop on Information Theory (CWIT), Ottawa, ON, Canada, May 2009.

H. Behroozi was with the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada. He is now with the Electrical Engineering Department, Sharif University of Technology, Tehran, Iran (e-mail: behroozi@sharif.edu).

F. Alajaji and T. Linder are with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: fady@mast.queensu.ca; linder@mast.queensu.ca).

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.2010.2040666